



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 118973

TO: Jennifer Graser
Location: 3b09 / 3c18
Wednesday, April 07, 2004
Art Unit: 1645
Phone: 272-0858
Serial Number: 09 / 844281

From: Jan Delaval
Location: Biotech-Chem Library
Rem 1A51
Phone: 272-2504

jan.delaval@uspto.gov

Search Notes

STIC-Biotech/ChemLib

118973

From: Hutzell, Paula
Sent: Wednesday, April 07, 2004 3:01 PM
To: Graser, Jennifer; STIC-Biotech/ChemLib
Subject: RE: rush search

approved

-----Original Message-----

From: Graser, Jennifer
Sent: Wednesday, April 07, 2004 12:26 PM
To: Hutzell, Paula
Subject: rush search
Importance: High

Hi Paula,

Could you please authorize the following rush search for an amendment which is due?

Thanks,
Jennifer

STIC:

Please search **SEQ ID NO: 1** from 09/844,281 in pending and commercial databases.

Thanks,
Jennifer Graser
REMSEN
EO3 B09/ 3C18 mailbox
AU 1645
272-0858

RECEIVED
APR - 7 2004
STIC

Searcher: Jan
Phone: 22504
Location: STIC
Date Picked Up: 4/7
Date Completed: 4/7
Searcher Prep/Review: co
Clerical: co
Online time: 1.0

TYPE OF SEARCH:

NA Sequences: ✓
AA Sequences: ✓
Structures: ✓
Bibliographic: ✓
Litigation: ✓
Full text: ✓
Patent Family: ✓
Other: ✓

VENDOR/COST (where applic.)

STN: ✓
DIALOG: ✓
Questel/Orbit: ✓
DRLink: ✓
Lexis/Nexis: ✓
Sequence Sys.: ✓
WWW/Internet: ✓
Other (specify): ✓

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:29:55 ; Search time 18 Seconds

(without alignments)
2409.691 Million cell updates/sec

Title: US-09-844-281-1

Perfect score: 4202

Sequence: 1 AGKFPDPVAGHWAEGSINY.....ITSEIGSQAVHVNLPNL 833

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4202	100.0	862	1 SLA2_BACAA	P94217 bacillus an
2	2833.5	67.4	874	1 SLAP_BACLI	P49052 bacillus li
3	719	17.1	814	1 SLA1_BACNA	Q90521 bacillus an
4	529	12.6	531	1 Y042_BACAN	Q90520 bacillus an
5	348	8.3	1176	1 SLAP_BACSH	P38537 bacillus sp
6	265.5	6.3	762	1 SLAP_ACEKI	P2258 acetogenium
7	252.5	6.0	941	1 GUN_BACS6	P19424 bacillus sp
8	230	5.5	1325	1 YDEK_ECOLI	P32051 escherichia
9	223.5	5.3	2003	1 YDBA_ECOLI	P33666 escherichia
10	219	5.2	1116	1 SLPH_ERECH	P38538 brevibacilli
11	217	5.2	1655	1 OMPB_RICCN	Q90543 r outer mem
12	211	5.0	1053	1 SLPM_BACBR	P06546 bacillus br
13	207.5	4.9	1645	1 OMPB_RICTY	P96989 r outer mem
14	207.5	4.9	2358	1 YEEJ_ECOLI	P76347 escherichia
15	207.5	4.9	2660	1 YEEJ_ECO57	Q88877 escherichia
16	206.5	4.9	1654	1 OMPB_RICRI	Q53047 r outer mem
17	206	4.9	1698	1 41_DROME	Q50477 drosophila
18	200	4.8	1643	1 OMPB_RICRP	Q3020 r outer mem
19	199.5	4.7	939	1 SLAP_CAMPE	P35827 campylobact
20	197.5	4.7	1608	1 HLYA_SERNA	P15320 serratia ma
21	195.5	4.7	1848	1 CSBA_CLOCL	P38058 clostridium
22	194.5	4.6	1656	1 OMPB_RICQA	Q06653 r outer mem
23	194.5	4.6	1902	1 P2P_LACPA	Q02470 lactobacilli
24	194	4.6	1300	1 120K_RICRI	P4914 rickettsia
25	193	4.6	917	1 SLAP_THETH	P38830 thermus the
26	189.5	4.5	1183	1 CNA_STAUA	Q33654 staphylococ
27	189	4.5	1953	1 BIGA_SALTY	P25927 salmonella
28	188.5	4.5	1276	1 PNP6_CHLFP	Q92899 chlamydia p
29	187.5	4.5	1861	1 APU_THETU	P38536 t amylopull
30	185.5	4.4	1577	1 HLYA_PROMI	P16466 proteus mir
31	183.5	4.4	1087	1 XYNX_CLOTM	P38535 clostridium
32	183	4.4	2249	1 OMPA_RICRI	P15921 rickettsia
33	183	4.4	4705	1 FAT2_DROME	Q50477 drosophila

ALIGNMENTS

RESULT 1
SLA2_BACAA STANDARD; PRT; 862 AA.
AC P94217; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE S-layer protein EAL precursor.
GN BAG OR BA0887.
OS Bacillus anthracis (strain Ames), and
OC Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094, 1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson J.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaiter J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sterne / 9131;
RX MEDLINE=97260111; PubMed=9106206;
RA Mesnage S., Tosi-Couture E., Mock M., Gounon P., Fouet A.,
RT "Molecular characterization of the Bacillus anthracis main S-layer
RT component: evidence that it is the major cell-associated antigen.";
RL Mol. Microbiol. 23:1147-1155(1997).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; A5017027; RAP24884.1; -
CC EMBL; X39724; CAA68063.1; -
CC TIGR; BA0887; -
CC InterPro; IPR001119; SLH.

Handwritten signature

```

DR Pfam; PF00395; SLH; 3.
KW Signal; Repeat; Cell wall; S-layer; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 862 S-LAYER PROTEIN EAL.
FT DOMAIN 34 76 SLH 1.
FT DOMAIN 95 136 SLH 2.
FT DOMAIN 157 197 SLH 3.
SQ SEQUENCE 862 AA; 91362 MW; CB16B20F62CCCA0 CRC64;

Query Match 100.0%; Score 4202; DB 1; Length 862;
Best Local Similarity 100.0%; Pred. No. 1.3e-175;
Matches 833; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGKSPDPVAGHWAGSINYLVDKGAITGKPDGTGPTESIDRASAAYFTKILNLPVDE 60
DB 30 AGKSPDPVAGHWAGSINYLVDKGAITGKPDGTGPTESIDRASAAYFTKILNLPVDE 89

QY 61 NAQPSFKDAKNWSSKYIAAEKAGVVGKDGKGFYPEGKIDRASFASMLVSAYNLKDKV 120
DB 90 NAQPSFKDAKNWSSKYIAAEKAGVVGKDGKGFYPEGKIDRASFASMLVSAYNLKDKV 149

QY 121 NGELVTTFFDLHDHWGEEKANILINLIGISVGTGGKWEPNKSVSRABAAQFIATDKKYGK 180
DB 150 NGELVTTFFDLHDHWGEEKANILINLIGISVGTGGKWEPNKSVSRABAAQFIATDKKYGK 209

QY 181 KDNAQAYVTDVKSPTKLTITGTLGDLKLSADDDVTEGDKAVAEASTDGTSAVVTLGGK 240
DB 210 KDNAQAYVTDVKSPTKLTITGTLGDLKLSADDDVTEGDKAVAEASTDGTSAVVTLGGK 269

QY 241 VAPNKDLTVKVNQSFVTKFYVEVKLAVEKLTTPDDDRAGQAIAFKLNDEKGNADVEYLN 300
DB 270 VAPNKDLTVKVNQSFVTKFYVEVKLAVEKLTTPDDDRAGQAIAFKLNDEKGNADVEYLN 329

QY 301 LANHDVKFVANNLDGSPANIPEGGEATSTGKLAIVGKQGVVQVTKGGGLTVSNTG 360
DB 330 LANHDVKFVANNLDGSPANIPEGGEATSTGKLAIVGKQGVVQVTKGGGLTVSNTG 389

QY 361 IITVKNLDPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGKASLNKIVA 420
DB 390 IITVKNLDPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGKASLNKIVA 449

QY 421 TIAGDKVVDPSGISIKSSNHGIIISVNNYITAEAGATLTIKVGDVTKDKVKVTTDS 480
DB 450 TIAGDKVVDPSGISIKSSNHGIIISVNNYITAEAGATLTIKVGDVTKDKVKVTTDS 509

QY 481 RKLVSVPKNDKLQVQNKTLPTVFTTDQYGDPPGANTAAIKVLPKTGVVAEGGLDW 540
DB 510 RKLVSVPKNDKLQVQNKTLPTVFTTDQYGDPPGANTAAIKVLPKTGVVAEGGLDW 569

QY 541 TTDSGIGTKTIGVNDVGEVTHFQNGATLSLYVNVTEGNAVAFKNFELSVKVGQY 600
DB 570 TTDSGIGTKTIGVNDVGEVTHFQNGATLSLYVNVTEGNAVAFKNFELSVKVGQY 629

QY 601 GQSPDTKLDLNVSTTVYQLSKYTSRDVYSDPENLEGYEVESKNLAVADAKIVGNKVVT 660
DB 630 GQSPDTKLDLNVSTTVYQLSKYTSRDVYSDPENLEGYEVESKNLAVADAKIVGNKVVT 689

QY 661 GKTPEGKVDIHLTKNGATAGATVEIVQETIAIKSNFNPQVTENVEKKNIGTVLELEK 720
DB 690 GKTPEGKVDIHLTKNGATAGATVEIVQETIAIKSNFNPQVTENVEKKNIGTVLELEK 749

QY 721 SNLDDIVKGINLTKETQHKVVRVVKSGAEGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS 780
DB 750 SNLDDIVKGINLTKETQHKVVRVVKSGAEGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS 809

QY 781 ALPNFKADLYDTLTKTKTGTLVFKVLKDKDVITSEIGSQAVHVNVLNPNL 833
DB 810 ALPNFKADLYDTLTKTKTGTLVFKVLKDKDVITSEIGSQAVHVNVLNPNL 862

RESULT 2
SLAP_BACLI STANDARD; PRT; 874 AA.

```

```

AC P49052;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE S-layer protein precursor (Surface layer protein).
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NM 105;
RC MEDLINE=97082965; PubMed=8964497;
RA Zhu X., McVeigh R.R., Malathi P., Ghosh B.K.;
RT "The complete nucleotide sequence of the Bacillus licheniformis NM105 S-layer-encoding gene."
RL Gene 173:189-194 (1996).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U38842; AAC44405.1; -.
DR PIR; JC4930.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
KW Signal; Cell wall; S-layer; Repeat.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 874 S-LAYER PROTEIN.
FT DOMAIN 33 93 SLH 1.
FT DOMAIN 94 155 SLH 2.
FT DOMAIN 156 217 SLH 3.
SQ SEQUENCE 874 AA; 92734 MW; EFADCA4FF27D32AF CRC64;

Query Match 67.4%; Score 2833.5; DB 1; Length 874;
Best Local Similarity 67.3%; Pred. No. 3.6e-116;
Matches 569; Conservative 98; Mismatches 156; Indels 23; Gaps 7;

QY 1 AGKSPDPVAGHWAGSINYLVDKGAITGKPDGTGPTESIDRASAAYFTKILNLPVDE 60
DB 30 AGKSPDPVAGHWAGSINYLVDKGAITGKPDGTGPTESIDRASAAYFTKILNLPVDE 89

QY 61 NAQPSFKDAKNWSSKYIAAEKAGVVGKDGKGFYPEGKIDRASFASMLVSAYNLKDKV 120
DB 90 NAQPSFKDAKNWSSKYIAAEKAGVVGKDGKGFYPEGKIDRASFASMLVSAYNLKDKV 149

QY 121 NGELVTTFFDLHDHWGEEKANILINLIGISVGTGGKWEPNKSVSRABAAQFIATDKKYGK 180
DB 150 DGLTVTKFDDLRHNGWEEKANILINLIGISVGTGGKWEPNKSVSRABAAQFIATDKKYK 209

QY 181 KDNAQAYVTDVKSPTKLTITGTLGDLKLSADDDVTEGDKAVAEASTDGTSAVVTLGGK 240
DB 210 PENSADKVTNVAATEPTQLTTLTGTGLNKLTAEVDTEGDKAVAEASTDGTSAVVTLGGK 269

QY 241 VAPNKDLTVKVNQSFVTKFYVEVKLAVEKLTTPDDDRAGQAIAFKLNDEKGNADVEYLN 300
DB 270 IAPNKELPVKVGNTFIVKVVYVEVKLVVEQLTFDDDRADQAVVFKLNDEKGNADIEYLD 329

QY 301 LANHDVKFVANNLDGSPANIPEGGEATSTGKLAIVGKQGVVQVTKGGGLTVSNTG 360
DB 330 IAGHDVKFVANNLDGTPANIPEGGTAEASTGKLAIVGAEKGVQVTKGGITVSNITG 389

QY 361 IITVKNLDPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGKASLNKIVA 420
DB 390 IIEVKNLDAEATAIKDVVFAVDTDKAG-VNYAKPLSGTDTLNSKTLVAGERAGIHKVA 448

```



```

Db 527 EFTAPVT--KYLDKDGKELKEQLEAKYVKNELV-----LNAAGCEAGNVTVVLTAKSGE 580
QY 571 GATLGLSVNVTEGNVAENFELVSKVGQYQSGPDTKLDLNVSTTVVQLSKYTSDRVYS 630
Db 581 KEAKATLAL--KAPGASKEFV-----RGLKELDKYVTEENQK 619
QY 631 PENLEGVEVSKNLAADAKIVGNKVVVTGKTGKVD-----IHLTKNGATA 678
Db 620 NAMTVSVLPVDAVGLVKGAEAAELKVTNTNKEGKGVDAATDAQVTVQNNSVITVGOGAKA 679
QY 679 GKA-TVEIVQETIAKSVNFPQVENFVKKINIGTVLEKSNLDDIVKGINL 732
Db 680 GETYKVVVLDGKLITHSFKVDTAPTAK-----GLAVEFTSLSKEVAPNADL 729

RESULT 4
YQ42 BACAN
ID YQ42 BACAN STANDARD; PRT; 531 AA.
AC Q9RMZ0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical cell-wall amidase PX02-42 precursor (EC 3.5.1.-).
GN PX02-42.
OS Bacillus anthracis.
OG Plasmid PX02.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA Kumano S., Menter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE N-
ACETYLURAMIDYL-L-ALANINE AMIDASE FAMILY 3.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF188935; AAF13647.1; -.
CC InterPro; IPR002508; Amidase_3.
CC InterPro; IPR001119; SLH.
CC Pfam; PF01520; Amidase_3; 1.
CC Pfam; PF00395; SLH; 3.
CC SMART; SM00546; Ami_3; 1.
CC PROSITE; PS01072; SLH DOMAIN; FALSE NEG.
KW Hypothetical protein; Hydrolase; Cell wall; Repeat; S-layer; Signal;
KW Plasmid.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 531 HYPOTHETICAL CELL-WALL AMIDASE PX02-42.
FT DOMAIN 33 72 SLH 1.
FT DOMAIN 91 132 SLH 2.
FT DOMAIN 153 193 SLH 3.
FT DOMAIN 400 521 AMIDASE.
FT SEQUENCE 531 AA; 59864 MW; 9A171DC4ED05CA78 CRC64;

Query Match 12.6%; Score 529; DB 1; Length 531;
Best Local Similarity 33.2%; Pred. No. 2.4e-16;
Matches 158; Conservative 68; Mismatches 184; Indels 66; Gaps 17;

QY 3 KSPDPVPAGHWAGSGINVLVDKGAITGKPDGTGVPTESIDRASAIVFTKIINLPVDENA 62
Db 31 KTTDVP--NWAQCSVNYLMKK-ALDGKPDGTGFSPEKIDRGAAKLMAVWLGQINKQA 87

```

```

QY 63 QPSFDKAKNWSKYIAAVEKAGVVGDKENFYPEGKIDRASPMASLVSNLXDKVNG 122
Db 88 KESFQDAKNHWSPIYIAAVEKAGVYGDGSGNFNFSKOIDRASMASMLVEAYKLNRIIG 147
QY 123 ELVTFPEDLLDHWGBEKANILNLGISVGTGSKWEPNKSVSRAEAAQFTALTDKYGGKD 182
Db 148 DJPTQPEDLKGEHGAALANALVALGISKGTGDKWKPNGIVTREAIVQPIAQDMK--KAD 205
QY 183 NQAQVYTD---VKVSEPTKLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVT-LG 238
Db 206 TSKRMVMNRHFTYTHQ---SLSSGVTNSQHAPOI-----IVVKEQRADGWIKIVTNG 256
QY 239 GKVAP--NKDLTVKVNQSFVT--KPVVEVKLAVEKLTFFDDRAGQAIAFKINDEKG-- 292
Db 257 DKWTPLYERETI---HSTFTTYPESHSSKVLGT-----HSPQTVT--VIEEGSW 303
QY 293 -----NADVEYLNLANHVDKVFVANNLDGSPANIPGGEATSTTGKLAIVGKQGD--YKVE 345
Db 304 IRIRTNASQWLDKQGLTLPEKQNNFLEKGTIIIDPHGGIDGGHKGKGYMNSPVVYDTA 363
QY 346 VQVTKREGGLTVSNTGIITVKNLDTFA-----SAIKNVFALDADNDGVVN----- 390
Db 364 VRQVKLPACKTFTALLTRDAYSPGKNATDSLKKEVEFAKKNKGDIFVSIHANGFNGA 423
QY 391 -----YGSKLGGKOPALNSQNLVVGKESLNLKLVATIGEDKVDVDPGSGISIKSSN 440
Db 424 HGTETEFYKAPTQKSNPYVNDSRILAEKIQ--KRLITALQTRDRGVKIGNLYVLREN 478

RESULT 5
SLAP BACSH
ID SLAP BACSH STANDARD; PRT; 1176 AA.
AC P38537;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Surface-layer 125 kDa protein precursor.
OS Bacillus sphaericus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1421;
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE=89371238; PubMed=2666389;
RA Bowditch R.D., Baumann P., Yousten A.A.;
RA "Cloning and sequencing of the gene encoding a 125-kilodalton
RA surface-layer protein from Bacillus sphaericus 2362 and of a related
RA cryptic gene.";
RA J. Bacteriol. 171:4178-4188(1989).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M28361; AAA50256.1; -.
CC PIR; A33856; A33856.
CC InterPro; IPR001119; SLH.
CC Pfam; PF00395; SLH; 3.
KW PROSITE; PS01072; SLH DOMAIN; 2.
KW Signal; Cell wall; S-layer; Repeat.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1176 SURFACE-LAYER 125 kDa PROTEIN.
FT DOMAIN 32 91 SLH 1.
FT DOMAIN 92 151 SLH 2.
FT DOMAIN 152 210 SLH 3.

```



```
Db 483 IGTLNMDATSSFDVTGIVNGFSGIVVNSGATLNSGCGF-----IGGNASGK 533
Qy 282 AIAFLNDEKGNADVEYLNLANHVDKVFVANNLGGSPANIEGGEATSTTG-----KL 333
Db 534 GIVNISTDSLWN-----LKTSSINAQLLVGLTGTGLMNTTGGIVKARDTQI 581
Qy 334 AVGIK-OGDYKVE-----VQVTRKGLVTSNTGIIITVKNLDT-----PASAI 374
Db 582 ALNDSKGDVRVDQNSLLETFFNNVYGTSGTGLTTLNNGTLNVEGGEVYLVGFPAVGT 641
Qy 375 KIV-----VFALDADNDGVNY-----GSKLSGK-- 398
Db 642 LNI GAAGHEAADAAGFTNATKVEFGGEGVFNHTNNSDAGQVDMITGDGDKGVI 701
Qy 399 ----DFALNSNLVVGKASINKLVATIAGE--DKVVDPSISIKSSNHGIIISVNNYIT 452
Db 702 HDAGHTVFNAGTYSG-KTLVNDGLLIIASHADGVTGMSSEVTIANPOTLDIL--AS 757
Qy 453 ABAAGEATLT-----IKGVDTKDKVKVT-----477
Db 758 TNSAGDYTLTINALKGDGLMRVQLSSDKMFGFTHATGTEFAGVAQLKDSFTTLERNTAA 817
Qy 478 ----TDSRKLVSVKANPKL-----QVQNKTLFVT-----FVTTD----- 509
Db 818 LTHAMLODSSENTSVKVGESIGGLAWNGGTIIIFDIPATLAEGYISVDLTVVGAGD 877
Qy 510 -----QVGDFFGANTAAIKVLPKTVVABGGLDVTTDS--G 545
Db 878 YTWKGRNVQVNGTGDVLIDVPKPWDPMPANNPLTLLLEHDD--SHVGQLVKAQTVIG 935
Qy 546 SIGTKTI-GVTGNDV-GEQVHF-QNGNGATLGLSYNVNT--EGNVAFKVF----- 591
Db 936 SGGSLTLDLQGEVEADKTLHIAQNGTVVABGDYGFELTAPNGLYVNYGLKALNIHG 995
Qy 592 ----ELVSKVQYQCSPTKL-----DLNVSTTVEYQLSKYSD-----RVYSD 631
Db 996 GQKTLAEGHAGYATADMSAKIGEGGLAINTVRQVSLNGQNDYQATVYQMGTLRTD 1055
Qy 632 PENLEGEVEKSLAVADAKIV--GNKVYV--TGKTPGV-----DIHLTKNGTAGK 680
Db 1056 ADGALG---NTRLEINSNAALVDLNGSTQVETFTGGQSTVLPKEGALTVNKGISQGE 1112
Qy 681 AT----VEIVQETIAIKSVNFK-PVQTFENFVEKINIGTVLEKSNL--DDIVKGNLT 733
Db 1113 LTGGNLNVGTGLAIEGLNARYNALTSPNAEVSVDNTQGLGRGNIANDGLLTKNVT 1172
Qy 734 KETQKVR--VVKSGAEQ-----GKLYLDENGDAVF-----NAGDVKL--G 770
Db 1173 GELARNSISGKIVSATARTDVELGDNSRFVGGQFNID-TGSALSVNEQKNGLDASVING 1231
Qy 771 DVTVSQSDSALPNPKADLYDTLTKYTDKGTLVFKLKD-----DVTISEI--GSQ 821
Db 1232 LLTISTERSWANTHSISGSDV----TKLGTGILTLNDSAAVQGTDIVGGEIAGFSD 1286
Qy 822 AV-----HVNVLNN 830
Db 1287 SAINMASQFHINHS 1301
```

RESULT 9

```
ID YDBA_ECOLI STANDARD; PRT; 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
GN YDBA OR B1401/B1405.
CS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
```

Query Match 5.3%; Score 223.5; DB 1; Length 2003;
Best Local Similarity 22.6%; Pred. No. 0.024;
Matches 173; Conservative 96; Mismatches 275; Indels 223; Gaps 38;

Qy 115 NLKDKVNGELVTTFFDLDPHNGEEKANILINL-GISVGTGGKWEFNKSVSRAEAQFIAL 173
Db 173 SLQDS-NGRKAT-----INLWQIDEANTVVALEGVSDGATKWQYNHN----- 214
Qy 174 TRKYGYKKNQAQYVTDVKNVSEPTKLTGTGLDKLSADDVTLGDKAVAIEASTDGTSA 233
Db 215 -----GELVITG-----DNATVNNNGTKTVDGKSTGTGE 243
Qy 234 VVTLGKGVAPNKDLTVKVNQSFVTKFVYEVKLAVEKLTFDDDRAGQAIAFKLNDEKGN 293
Db 244 INGNNGKVIQDGLDVSQGGHGI-----DITGDSATV---DNKGT 280
Qy 294 ADVEYLNLANHVDKVFVANNLGGSPANIEGGEATSTTGLAVGIKQGDYKVEVQVTKRG 353

```
Db 281 MTVT-----DPESMGIGIDGDKAIVNNEGSEITNGGTGTQINGDD----- 321
Qy 354 LTVSNTGIIITVKNLTPASAIKNVFPALDADNDGVVYNGS-KLSGKDFALNSQNLVWGEK 412
Db 322 ATANNNGKTTVDGKSTGTETI-----NGNNGKVIQDGLDVSOGGHHID-----ITGDS 370
Qy 413 ASINKLVATIAGEDKVVDPGSIKSSNHGIIISVNNVITABAAEATLT-----TKV 465
Db 371 ATVDN-----KGTMTVTDESIGIOVD--GDOAVNN-----EGESAITNGGTGTQING 417
Qy 466 GVTVDKVKVTTDSKLVSVKXANPKLOVQNKTLPTV-----FVITDQVG-DPFGAN 518
Db 418 DDATAANNNGKTTVDGKSTGTETIAGNNGKVIQDGLDVSOGGHHIDITGDSATVDNKGTM 477
Qy 519 TAAKEVLPTKTVGAEGGLDV-----TTDSGSIKTGTIG--VTGNDVGEVTHFQNGN 570
Db 478 TVTDPESI--GIQIDGDAIVNNEGSEITNGGTGTQINGNDATANNNGKTTVDGKST 534
Qy 571 GATL-----GSLIV-----NUTE-GNVAFKMFELVSKVQY-----GOSP 604
Db 535 GTKIAGNIGIVNLDGSLTITGGAHGVENTGDNVTNNKGDIVVSDTGSIGVILNGEGATV 594
Qy 605 DTKLNLVS-----TTVEYOLSKYTSDRV--YSDPENLEGYEYESKNLAVADAKIVG 654
Db 595 SNTGDNVNSNEATGFSITTSNGSKVSLAGSVQGVDFSTGVYDLGN--NSVTLAAKDLKVVG 653
Qy 655 -----NKVVVTG-----KTPGK-----VDHLTKNGATA-GKATV 683
Db 654 QKATGINVSGDANTVITGNVLVDKDTADNAAEYFPDPFSGVINGVSDNNVTLDPGLTV 713
Qy 684 EIVQETIAIKSVNF--KPVQTFNVF---EKKINIGTVLEL--EKNLDDIVKGINLTKE 735
Db 714 VSSEVTSRQSLNFDGSAETSLGIVIGDNTVNMGGELICEKNALADGSGQVTSLRG 773
Qy 736 TOHKRVYKSGABQKLYLDRNGDAVFNAGDVKLGDVTVTSQSDSAL 782
Db 774 YSVTSVIVVSG--ESSVYL--NGDTTI-SGEFFLGFAGVIRVQDKAL 815

RESULT 10
SLPH_BRECH STANDARD; PRT; 1116 AA.
AC P38538;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Surface layer protein precursor (Hexagonal wall protein) (HWP).
OS Brevibacillus choshinensis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
CX NCBI_TaxID=54911;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 54-68.
RC STRAIN=HPD31.
RX MEDLINE=90170842; PubMed=2307650;
RA Ebisu S., Tsuboi A., Takagi H., Naruse Y., Yamagata H., Tsukagoshi N.,
RA Uda S.;
RT "Conserved structures of cell wall protein genes among protein-
RT producing Bacillus brevis strains.";
RL J. Bacteriol. 172:1312-1320(1990).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90050; BAA14103.1; -.
```

```
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 2.
DR PROSITE; PS01072; SLH DOMAIN; 2.
KW Signal; Cell wall; S-layer; Repeat.
FT SIGNAL 1 53 SURFACE LAYER PROTEIN.
FT CHAIN 54 1116 SLH 1.
FT DOMAIN 57 120 SLH 2.
FT DOMAIN 121 171 SLH 3.
FT DOMAIN 172 231 SLH 3.
SQ SEQUENCE 1116 AA; 123397 MW; 86D58D37AC72546F CRC64;

Query Match 5.2%; Score 219; DB 1; Length 1116;
Best Local Similarity 19.2%; Pred. No. 0.018;
Matches 202; Conservative 142; Mismatches 347; Indels 362; Gaps 47;

Qy 15 EGSINYLVDKATGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ-----PSFKDA 69
Db 68 ETKVRLLEALGLVAGYNGDFGADKTLITRAEFTALIVRAGL--EQAKLAQNFVTTVDV 125
Qy 70 KNI-WSSKYIAAVEKAGVKGDKGFENFYPEGKIDRASAFSMLYSAYNLKDKVNGELVTTF 129
Db 126 RSTDFAGFVNVASGEIIVKGFDPKSPQNVTYAEAVTMIVRALGYEFPVSRGV----- 180
Qy 129 ELLDHGHEKANKILNLIGISVGTGKWEKPNKSVSRAEAAQFIALTDKYKKDNAQ--- 185
Db 181 -----W-----PNSMISKSELNIA-----KGINPNMQQFAATIFML--DNALRVK 221
Qy 186 -----AYVDVK--VSEPTKLT-----VAIEASTDGTSAVVTLGK-----VAP 243
Db 222 LMEQIEGTDIRLNVITDETLTKLVYVRDMWAHEKGNNSDELPLTVNVPAIGLSLK 281
Qy 211 ADDVTLEGDKA-----VAIEASTDGTSAVVTLGK-----VAP 243
Db 282 ANEVTLNGKADLGSNTTYKVAEGINPNAFDGQKVQVWIKDDRENVIVWMEGSEDEVDVM 341
Qy 244 NKDLTVKVKQNSFVTKFVVEVKLAVKLTDFDDDRAGQATAFKLNDE-----KGNAD 295
Db 342 DRVSAIYLGKAFATDDIVKDLKSDLDVVKIEMD--GSEKSYRLTEDTKITNYTFRNDP 399
Qy 296 VEYLNLANHD-----VKFVAN-NLDGSPANIFEGEATSTTGKLVAGIKQGDYKVEQVT 349
Db 400 VDALSKIYKNDNTPGVKVLNNDNNEVAYLHIID-----DQITDKSVKGVKYG----- 446
Qy 350 KRGGLTVSNTGIIITVKNLDTPASAIKNVFPALDADNDGVVNY--GSKLS-----GKDA 401
Db 447 -----SKVIDADKKITNLDNSKFSLEDDQDEGKDFL 481
Qy 402 LNSQNLVVGKASLNKLVIATIGEDKVVDPGSIKSSNHGIIISVNNVITABAAEATLT 461
Db 482 V-----FLDGQPAKL-----GDLKESDVSVYVYADGDKYLVFANRNVAEGKVEKV 529
Qy 462 T-----IKVGDVT-----KDVKEKVTDSREKLVSKANPKLOVQNK 500
Db 530 SRNKTDIRLTVGGKTYKVPDASYSENANKDKV-KVNSDDLILNLDGEEVKLL----- 583
Qy 501 LPVTEVTTDQYGDPPGANTAAIKVLPKTVGVAEGGLDVVTTDSGSIKTGTIGVTGN--- 557
Db 584 -----DPSG-----RVRIETKDAIDRRKPLAIITK-----GATYNSK 617
Qy 558 DVGEGTVHFQNGNGATLG-----SLY-----VNVTEGNVAFKNFE--LVSKVQYQSGSDTK 607
Db 618 DTYDFTVMTQKGTQIVSLDQKDIYDRYGVNYDKNRQAFKDLVELLOKVKVVEDSA 677
Qy 608 LQLNVSTTVVEYQL-----SKYTSDRVSDPENLEG-YEYESKNLAV 647
Db 678 TDAQTVLLEVNFDGSGVEVDKVLDSLKYSEKSTWKLADDDDDVVDGVEYTDK---T 734
Qy 648 ADAKIVGNKVVVTKTPGKVDIHLTKNGATA-----GKATVEIV-- 686
Db 735 AVFKMTGDLTPATGTRGEL-----KNAGTAKFDKVAKSKDLKVVYSVDEDKGEVCAIFV 789
Qy 687 -----QETIAIKSVNFQVQTFNF-----VEKKINIG 713
```

```

Db 790 VDSGLGGHGHQFQWVKQYGTASKQDITITVTKDGDVTEKEYKLDGDDDLKVDQOIRG 849
QY 714 TVLELEKSN-----LDDIVKGIN-----LTKETQKRVVKSAGQKGLYLD 755
Db 850 DVISFTLNSDGEVIVDDVVEVNNHIDNTASKSATLMPEDERQ-----KAGID--KLVA 903
QY 756 RNCDAVFNAGDVKLGVTVSQTSDSALPNFKADLYDTLT--TKYTDKG----- 801
Db 904 RYDEVGNTISLNYAGDKTKQYTKASTAF-IDVYDGLGIDGVDEGVYVWIDSADIDG 962
QY 802 ---TLVFKVLKDKDVITSEIGSQAVHVNINNP 831
Db 963 TRDYVLVWSSDDEIRTOHISTKAV-TDLNKP 994

RESULT 11
OMPB_RICCN STANDARD; PRT; 1655 AA.
AC Q9KK98; Q9XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scd5) (rOMP)
DE (OMP B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB OR R1085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart F., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [2]
RP SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein rOMP (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN [3]
RP SEQUENCE OF 353-1655 FROM N.A.
RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a s-
CC layer with hexagonal symmetry (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; AB008659; AAL03623.1; -
DR EMBL; AF123721; AAF34124.1; -
DR EMBL; AF123726; AAF34129.1; -
DR EMBL; AF149110; AAD39533.1; -
DR PIR; E97835; E97835.
DR InterPro; IPR006315; Autotransport.
DR Pfam; PF03797; Autotransporter.
DR TIGRFAMs; TIGR01414; autotrans_barl.2.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1334
FT VARIAT 61 61
FT VARIAT 75 75
FT VARIAT 78 78
FT VARIAT 251 251
FT VARIAT 413 413
FT VARIAT 959 959
FT VARIAT 988 988
FT VARIAT 1139 1139
FT CONFLICT 353 354
FT CONFLICT 776 776
FT CONFLICT 1159 1159
FT CONFLICT 1177 1177
FT CONFLICT 1492 1492
SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

Query Match 5.2%; Score 217; DB 1; Length 1655;
Best Local Similarity 20.5%; Pred. No. 0.035;
Matches 183; Conservative 123; Mismatches 336; Indels 252; Gaps 45;

QY 104 ASFASMLVSAYNLKDKNGELVTTFEDLLDHGWEKANKILNLGIVSGTGKWEPNKSVS 163
Db 26 ASFASMGAAIQQRNTNVAITVDGVGFDQTPANVAVPLNAVITAG-----VNKGIT 81
QY 164 -RAEAAQFIAL---TDKKYG---KKDNAQAYVTDV-----KVSEPTKLTITGTGLD 207
Db 82 LNTFAGSNGFLFNTANNLDVTVEEDTLGPIIVNANNHNLMLNAGKLTITGCGIT 141
QY 208 KL-----SADVTLEGKVAIE-----AST-----DG 230
Db 142 NVQAAATKANNVVAQVNGAAIDNNDLQGVGRIDCGAAASTLVFNLANPTTKAPILG 201
QY 231 TSAVVTLLGKVAPN-KDLTVKVKQSFVTKFVYEVKVLAVEKLTFFDDDRAGQAIAFLND 289
Db 202 DNAVINGANGTLVNTGFIKVSSEKSPAT-----VNVINIGD---GGIMENIDA 248
QY 290 EKNADVEYLNLANHDKVFNANNLDGSPANIFEGGEATST---TGKLAVGIKQ---GD 341
Db 249 D-----NVNTLNLQANGATITFNGTGTGRVLVLSKNAATDPNVTGSLGNLKGIIENPT 304
QY 342 YKVEVQVTKRGGLTVSNIGIITVKNLDPASAIKNVVFALDADNDGVVNYGSKLSGKQFA 401
Db 305 VAVNGQLKANAG---ANAAVIGTNV-----CAGRAAGFVSVSDNGKVATIDGVYAKDMV 356
QY 402 LNSQNLV-----VGE-----KASLNKLVIATIGEDKVVDPGSIK- 437
Db 357 IQSANAVGVNFRHIVDVGTDGTTAFKTAASKVAITQNSNFGTTFDFGNLAAQIIVPNTMT 416
QY 438 -----SSNHG-----LISVNVNVIATAEAGEATLTIKVGDV 468
Db 417 LINGNFTGASNPNTAGVITFDANGTLASADANVANVTNITALEASGAGVQLS-GTH 475
QY 469 TKDVK-----FKVTTDSRKLVSVKANPKLQVQVQNKLPVFTVTTDQYDPPGANTAA 521
Db 476 AAEELRLGNAGSVFKLADGT--VINGKVNQAL-----VGALAAGATIT 516
QY 522 IKEVLPKGVVABEG---LDVTTDSGICTKIGTVGNDVGE--GTVHQNGNGATLG 575
Db 517 LDGSATITGDIAGGAALAOQITLANDATKTLTG-GANLIGANGGHINQ-ANG---G 571
QY 576 SLVNVNVTGNAFKNFELVSKVGQYQSGPDTKLDLNVSTTVEYQLSKYTSRVSDDPNL 635

```

Db 572 TIKLTSTONNIVV-DPDLAIATDQTGVWDASSLTNAQLTINGKIG-----TVGANNKTL 625

Qy 636 EGYEVESKNLAVADAKIVGKVVVTKTPGKD-----IHLTKNGATAGATVEIV-- 686

Db 626 GQENIGSSKTVISDGVAINELVIGNN--GAVQFAHTVLTIRITNAAGQKLIENPVN 683

Qy 687 QETIAIKSNVF-----KPVOTENFVEK-KINIGTVLEKSNLDDIVKGINL--TKETQHK 739

Db 684 NNTTLATGTLGSAATPLAEINFGSGAANVDTVL-----NVGKGVNLYATNITTTD 735

Qy 740 VRVVK-----SGAQGKLY--LDRN-----GDAVENAGDVKLGDVTVS 775

Db 736 NVGSIIFNAGGNIVSGTVGGQGNKFNVALDNGTTVKFIGNATFN-----GNTTIA 789

Qy 776 QTSDSAL-PNFKADLYDTLTTKYTKDGLTVFKVKKDKOVITSEIGSQAVHNVN 828

Db 790 ANSTLQIGGNTADVFASA-----DGTGIVEFV-NTGPIITVLNKAAPVNAL 836

RESULT 12

SLPM_BACBR STANDARD; PRT; 1053 AA.

AC P06546;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Middle cell wall protein precursor (MWP).

OS Bacillus brevis (Brevibacillus brevis).

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.

CX NCBI_TaxID=1393;

RN SEQUENCE FROM N.A.

RP STRAIN=47;

RC MEDLINE=88115203; PubMed=2828336;

RX Tsukagoshi N., Uchihashi R., Adachi T., Sasaki T., Hayakawa S., Yamagata H., Tsukagoshi N., Uchida S.;

RA "Characterization of the genes for the hexagonally arranged surface layer proteins in protein-producing Bacillus brevis 47: complete nucleotide sequence of the middle wall protein gene.";

RT J. Bacteriol. 170:935-945 (1988).

RN [2]

RP SEQUENCE OF 1-199 FROM N.A.

RC STRAIN=47;

RX MEDLINE=87137282; PubMed=3023027;

RX Yamagata H., Adachi T., Tsukagoshi A., Takao M., Sasaki T., Tsukagoshi N., Uchida S.;

RA "Cloning and characterization of the 5' region of the cell wall protein gene operon in Bacillus brevis 47.";

RT J. Bacteriol. 169:1239-1245 (1987).

RN [3]

RP SEQUENCE OF 676-1053 FROM N.A.

RC STRAIN=47;

RX MEDLINE=87008404; PubMed=2428810;

RX Tsukagoshi A., Uchihashi R., Tabata R., Takahashi Y., Hashiba H., Sasaki T., Yamagata H., Tsukagoshi N., Uchida S.;

RA "Characterization of the genes coding for two major cell wall proteins from protein-producing Bacillus brevis 47: complete nucleotide sequence of the outer wall protein gene.";

RT J. Bacteriol. 168:365-373 (1986).

RN [4]

RP SEQUENCE OF 1-50 FROM N.A.

RC STRAIN=47;

RX MEDLINE=900781123; PubMed=2512285;

RX Tsukagoshi A., Uchihashi R., Engelhardt H., Hattori H., Shimizu S., Tsukagoshi N., Uchida S.;

RA "In vitro reconstitution of a hexagonal array with a surface layer protein synthesized by Bacillus subtilis harboring the surface layer protein gene from Bacillus brevis 47.";

RT J. Bacteriol. 171:674-6752 (1989).

RN [5]

RP FUNCTION: The middle wall protein binds to peptidoglycan and to the outer cell wall protein.

CC -!- SUSUNIT: THE MIDDLE CELL WALL LAYER IS COMPOSED OF SUBUNITS OF THE MIDDLE CELL WALL PROTEIN. THESE PROTEINS FORM A HEXAGONAL

CC ARRAY WITH A LATTICE CONSTANT OF 14.5 NM IN THE MIDDLE CELL WALL LAYERS.

CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.

CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M15364; AAA87321.1; -

DR EMBL; M14238; AAA22372.1; -

DR EMBL; M19115; AAA22760.1; -

DR EMBL; M31828; AAA22619.1; -

DR PIR; A28555; A28555.

DR InterPro; IPR001119; SLH.

DR Pfam; PF00395; SLH; 2.

DR PROSITE; PS01072; SLH_DOMAIN; 2.

KW Cell wall; S-layer; Signal; Repeat.

FT SIGNAL 1 23 MIDDLE CELL WALL PROTEIN.

FT CHAIN 24 1053

FT DOMAIN 29 92 SLH 1.

FT DOMAIN 93 143 SLH 2.

FT DOMAIN 144 203 SLH 3.

FT DOMAIN 153 AA; 117146 MW; DB421318BD9D5E4F CRC64;

SQ SEQUENCE 1053 AA; 117146 MW; DB421318BD9D5E4F CRC64;

Query Match 5.0%; Score 211; DB 1; Length 1053;

Best Local Similarity 19.9%; Pred. No. 0.036;

Matches 193; Conservative 138; Mismatches 321; Indels 320; Gaps 45;

Qy 15 EGSNYLVKCAITKDPDGYGTPESIDRASAAVIFKILNLPVDENAQ-----PSFKDA 69

Db 40 ETKVRLALGLVAGYNGEYGVNDKTTIRAEFAFLVVRAGL--EQAKLAQFNTYDV 97

Qy 70 KNI-WSSKYAAVEKAGVVGKDGKGFYKPKQVQVYAEAVTMVVALGVPSVKGV----- 128

Db 98 KSTDFAGFNVASGEIIVKGFDPKPKQVQVYAEAVTMVVALGVPSVKGV----- 152

Qy 129 ELLDHWGE---EKANILINIGISVGTGGKWEKPNKSVRAEAAQFIATDKYKKNDA- 184

Db 153 -----WPNMSISKASEL-NIARSITT-----PNAATRGDIFKOL-----DNAL 190

Qy 185 -----QAVYTDVK--VSEPTKLT-----LTGTGLD 207

Db 191 RVDLMEQVEFGTDIRHEITKELTLTKYLKVTVRDMEWAQEAQNDSEDLPLVTNVPALGLG 250

Qy 208 KLSADDVTLEGKAVAIEAST-----DGTSAVVTLGKVA-----PNKDLTVKVNQSFVT 258

Db 251 KIKANEVTLNG-KDAGIGNTTKYVADGINANDPDGQHVQVWIKDDKEDVIWMSEGSTDQE 309

Qy 259 KFYVEVKLAIEKLTFDDDRAGQAIAFKLNDEKNADVEYLINLHNDVHVKFVANNLDGSPA 318

Db 310 VMIDRVGEFTLKGTFEDPK-----DLNSDL----- 336

Qy 319 NIFEGEATSTTGKLVAGIKQGVKVEQVQVTKRGGLTVSNVTGI-ITVKNLDTPSAIAKV 377

Db 337 -----ADLKLELDASEKSYRFNKNTKVTYNTFRNDFPDVGLKEI 375

Qy 378 VFALDADN-DGVVNYGSKLSCKDFALNSQNLV-----VGEKASLNKLVAIAGEDKVVDPG 432

Db 376 I-----KNADGGFTFGAKV-----VLONNNEIAYIHIDDOSMMK-----EEGVKYG 419

Qy 433 SISIKSSNHGISVYNNYITAAAGAEATLTIKGVGVTKVQKFTVTSRKLVSUKVANDPK 492

Db 420 S-----EVISKIDTDKKKITN-----RDNDK 440

Qy 433 LQVQNKTLPTVFTVTDQYGGPFGANTAAIE-----VLPTKTVGVAEGG 536

Db 441 FNDLDGKEGKDFLVF-LNGKPF--AKFSDLKEGMVSVIYADGDEDKLLLVFATDTVVEGK 497

587 AFKNFELVSKVGQSPDKLNLVSTTVEYQSKYSDRVVSDPENLEGVEVESKNLA 646
 584 -LVDFDLVDTDTGVDASSLNNQTLFNGSIG-----TIGANTKTLGRFNVGSSKTI 637
 647 VADAKIVGNKVVTGKTPGKVDIHLTK-----NGATAGKATV---EIVQETIAIKS 694
 638 LNAAGDVAINELVM--ENDGSV--HLTHNTYLIITKTINAANQGIIVAAADPINTDTALADG 693
 695 VNF---KPVOTENFVEKKNIGVLELEKSNLDDVKGINL-----TKETQHKV----- 740
 694 TNLGSABSPISNTHFATKAANGDSILH-----IGKGVNLVANNITTDANVGSLSHFR 745
 741 -----RVYKS--GAQQG---KLXLDN-----GDAVFNAGDVKLGDVTVSQTSDBALFN 784
 746 SGGTSIVSGTVGGQGLKNNLIDNGTIVKFLGDIITFNGG--TKIEGKILQISSNYIID 804
 785 F--KADLYDTLTKYTDKGLV-----FKVLKD-----KDVITSEIGSQAVH 824
 805 HIESADNTGTFVNTDPIITVTLNKQGYFGLVKQVWVGPNIAFNEIGNGVAH 859

RESULT 14
 YEEJ_ECOLI STANDARD; PRT; 2358 AA.
 AC P76347; P94750;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yeeJ.
 GN YEEJ OR B1978.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayaishi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Saeki G., Seki Y., Sivasubramanian S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 CC -!- SIMILARITY: Contains 13 Big-1 domains.
 CC -!- SIMILARITY: Belongs to the intimin/invasin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AE000289; AAC75042.1; ALT_INIT.
 DR EMBL; D90837; BAA15800.1;
 DR EMBL; D90836; BAA15799.1; ALT_INIT.
 DR EcoGene; EG13378; yeeJ.

DR InterPro; IPR003344; Big-1.
 DR InterPro; IPR003535; Intimin.
 DR InterPro; IPR008964; Invasin_intimin.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR000601; PKD.
 DR Pfam; PF02369; Big-1; 13.
 DR PRINTS; PR01369; INTIMIN.
 DR SMART; SM00634; BID_1; 13.
 DR SMART; SM00257; LysM; 1.
 DR SMART; SM00089; PKD; 6.
 KW Hypothetical protein; Repeat; Complete proteome.
 FT DOMAIN 738 834
 FT BIG-1 1.
 FT DOMAIN 840 931
 FT BIG-1 2.
 FT DOMAIN 932 1033
 FT BIG-1 3.
 FT DOMAIN 1042 1137
 FT BIG-1 4.
 FT DOMAIN 1146 1237
 FT BIG-1 5.
 FT DOMAIN 1246 1350
 FT BIG-1 6.
 FT DOMAIN 1351 1448
 FT BIG-1 7.
 FT DOMAIN 1449 1553
 FT BIG-1 8.
 FT DOMAIN 1554 1655
 FT BIG-1 9.
 FT DOMAIN 1656 1754
 FT BIG-1 10.
 FT DOMAIN 1755 1853
 FT BIG-1 11.
 FT DOMAIN 1853 1950
 FT BIG-1 12.
 FT DOMAIN 1952 2053
 FT BIG-1 13.
 FT CONFLICT 105 105 S -> G (IN REF. 2).
 SQ SEQUENCE 2358 AA; 248599 MW; 232249750BF631ED CRC64;
 Query Match 4.9%; Score 207.5; DB 1; Length 2358;
 Best Local Similarity 20.2%; Pred. No. 0.14; 390; Indels 225; Gaps 39;
 Matches 188; Conservative 126; Mismatches 390; Indels 225; Gaps 39;
 QY 17 SINYLVDRKA-----ITGKPDGTGYPTESIDRAGAAVFTKILNLPVDENAPQSPKDAKN 71
 DB 1236 TVHFIGDTAAAKIIEIAPVPDSIIAGTP--QNSSGSGVITATV---VDNNGFPVKGVTVN 1289
 QY 72 IWSSKVIIAVEKAG--VVKGDGKENF-----YPEGKIDRASFAFMLVSA 113
 DB 1290 FTSNAATAEMTNGGOAVTNEQKATVYTNTRSSIESGARDPTVVASLENGS--STLSTS 1347
 QY 114 YN-----LKDQVN--GELVTTPEDLIDHWGEEKANILINLIGSVGTG---- 153
 DB 1348 INVNADASTAHLTLQALFDIVSAGETTSLEIVKDNVGNVGPQOEVLTVSPSEGVTPS 1407
 QY 154 -----GKWFNPKSVSRARAAQFIATLDKKYKQKONAAQAYTVDVKVSEPTKLTGT 203
 DB 1408 NNALTYTNHNGNFYASFTATKAGVYQLTATLE---NGDSMOQTYYVFNVANASITLAA 1463
 QY 204 TGLDKLSADDVTLGDKKAVAIEASTDG-----TSAVVTL-----GGKVAPNKD 246
 DB 1464 S--KDPVIADNDL--TTLTATVADTEGNAIANTEVFTFLPEDVKANFTLSDGKVIITDAE 1520
 QY 247 LTVKV-----KNQSFVTKFVEYVKKLAVEKLTFFDDDRAGQAIAP-- 285
 DB 1521 GKAKVTLKGTKAGATVVTASMTGGKSEQLVWNFIADTLTAQVNLNVNEDNF--ANNVGMTR 1580
 QY 286 ---KLNDKGNADVVELNLANDHVKFVANNLDGSPANIF-----EGGEA--TSTTGKLA 335
 DB 1581 LGATVTDGNGNP-----LANEAVFTTL-----PADVSASFLLGGGSAITDINGKA 1628
 QY 336 ---GKQGDYKVEYQVTKRGGLTVSNTGIIIVKNLDTPASAKNVVFAALDANDGVNNG 392
 DB 1629 TLGKTKSGTYPTVTVSNVNYG---VSDT-----KQVTLIADA---GTAKLA 1667
 QY 393 SKLSGKDFALNSQ-----NLVVGKASLNKLAVATIAGEDKVDGPGSISIKS 438
 DB 1668 SLTSVYSFVSVSTEGATMTASTVDANGNPVEIKVNFRTSTVLS-----STSVET 1718
 QY 439 SNHGIISVNNYITAEAGEATLTKVGD-----VTQDKVKFKVTTDSRKLVSXKANPDKL 493
 DB 1719 DGRGFREIL---VTSEVGLKTVSASLADKPTEVISRLINAGADVNSATITSLIEPEGQV 1775
 QY 494 QVVQNKTLPTVFTTDDYQDPPGANTAAKVLPTKVTVVAEGGLDVVTTDSSGIGTKTIG 553

Db 1776 MVAQD--VAVKAVNDQFNPVAPQVFPFAESSQWILSON-----TVSTNTQGVAEVT 1828

Qy 554 VTGNDVGEVTHFQNGGATLGLSVNVTGNAVAFKNFELSVKGYGQSPDKLDLNV 613

Db 1829 MTPERNGSYVVKASLPNGASLEKQLEAIDEKLTLTASSPLIGVYAPTGTATLTATLSANG 1888

Qy 614 TTVEYQLSKYTSRVYSDPENLEGEYVESKNLAVADAKIVGNK--VVTGKTPGKVDTH 670

Db 1889 TPVEGQVINF--VTEGATLGGKVRITNSGQAPVVLTSNKVGYTVTASPHNGVTIQ 1945

Qy 671 -----LTKGATAGKATVEIVQSTIAIKSVNFKPVQ-----TENFVEKIMINGVLEL 718

Db 1946 TQTTVKVTGNSSTAWASFADPSTIAATNTDLSTLKATVEDGSGNLIIEGL-----TVYFA 2001

Qy 719 EKSNDLIVKGINTKETQHKVRVVKSGAGCKLYLDRNGDAVFNAGDVKLGVTV--- 774

Db 2002 LKSGSATLSTAVTDQNGIATTSVK-GAMTGSVTV---SAVTTAGGMVTDTILVAGP 2056

Qy 775 SQTSDSALPNFKADLYDTLTTKYDKGTL 803

Db 2057 ADTSQVLKSNRS---SLKGDYDTSDEL 2081

RESULT 15

ID YEEJ ECO57 STANDARD; PRT; 2660 AA.

AC Q8X8V7; Q8X2B9; Q8X2C0;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein YeeJ.

GN Z3135 OR ECS2775/ECS2776.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxId=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,

RA Grobeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamocousis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"

RL Nature 409:529-533 (2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12.;"

RL DNA Res. 8:11-22 (2001).

CC -1- SIMILARITY: Contains 16 Big-1 domains.

CC -1- SIMILARITY: Belongs to the intimin/invasin family.

CC -1- CAUTION: Ref.2 sequence differs from that shown due to a

CC frameshift in position 1315.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AEO05423; AGS57041.1; ..

DR EMBL; AP002559; BAB36198.1; ALT_FRAME.

DR EMBL; AP002559; BAB36199.1; ALT_FRAME.

DR InterPro; IPR003344; BIG-1.

DR InterPro; IPR003335; Intimin.

DR InterPro; IPR008964; Invasin_intimin.

DR InterPro; IPR000601; PKD.

DR Pfam; PF02369; Big_1; 16.

DR PRINTS; PR01369; INTIMIN.

DR SMART; SM00634; BID_1; 16.

DR SMART; SM00089; PKD_8.

KW Hypothetical protein; Repeat; Complete proteome.

FT DOMAIN 738 834 BIG-1 1.

FT DOMAIN 840 929 BIG-1 2.

FT DOMAIN 931 1033 BIG-1 3.

FT DOMAIN 1042 1132 BIG-1 4.

FT DOMAIN 1134 1236 BIG-1 5.

FT DOMAIN 1245 1335 BIG-1 6.

FT DOMAIN 1337 1439 BIG-1 7.

FT DOMAIN 1448 1539 BIG-1 8.

FT DOMAIN 1548 1652 BIG-1 9.

FT DOMAIN 1653 1750 BIG-1 10.

FT DOMAIN 1751 1855 BIG-1 11.

FT DOMAIN 1856 1957 BIG-1 12.

FT DOMAIN 1963 2056 BIG-1 13.

FT DOMAIN 2065 2156 BIG-1 14.

FT DOMAIN 2157 2252 BIG-1 15.

FT DOMAIN 2254 2355 BIG-1 16.

FT SEQUENCE 2660 AA; 01EB92A08F5C09D2 CRC64;

Query Match 4.9%; Score 207.5; DB 1; Length 2660;

Best Local Similarity 20.3%; Pred. No. 0.17;

Matches 197; Conservative 134; Mismatches 387; Indels 253; Gaps 46;

Qy 17 SINLYVDKGA-----ITGKPDGTGYGTSIDRASAIVFTKILNLPVDENAPSPFKDKN 71

Db 1538 TVHFIGDTAAAKIIEITPVPDSIIAGTP--QNSSGSVITATV---VNNNGFPVKGVTN 1591

Qy 72 IWSKYIAAVERAG--VVKGDGKENF-----YPEGKIDIRASFASMLVSA 113

Db 1592 FTSNANTAETNGGQAVTNEQKATVYTNTRSS-ESGARPDVTEASLENGS--STLSTS 1649

Qy 114 YNLKQKNGELVTTFDLLDHGE-EKANILNLGISVGTG-GKHEPNKSVSRRAAQ-- 169

Db 1650 INVNADASTAHLTLQALFDTVSAGDTNLYIEVDKNGYNGVPOQEVTLVSFSEGVTPS 1709

Qy 170 -----FIALTOKYK-----KKDNAQAVVT-----DV 191

Db 1710 NNALYTNHDGNFYASFATKAGVYQVVTATLENGDSMQQTVVVPVWANAETSLASKDP 1769

Qy 192 KVSEPTKL-TLTGTGLD---KLSADDVTL-----EGDKAVAI EASTGTSA 233

Db 1770 VIANNNDLTATVADTEGNAIANSEVFTLPEDVRANFTLGDGKKV---TDTEG-KA 1825

Qy 234 VVTLLGKVAIPNKDLTVKV---KNQSFVTKFVVEVKLAVEKLTFFDDDRAGQAIAP----- 285

Db 1826 KVTLLGKTGAHTVTASMAKGSEQLVNVFIADTLTAQVNLNVTEDFIANNVGMTRLOA 1885

Qy 286 KLNDERGNADVEYLNLANDHVKFVANNLDGSPANIF-----EGGEA-TSTTGKLAIV--- 335

Db 1886 TVTDGNGNP-----LANEAVITFL-----PADVSASFLLGQGSAITDINGKAETLS 1933

Qy 336 GIQGDYKVEVQVTKGGLTVSNTGIIITVKNLDTPSAIAKNVVFALDADNDGVVNYGSKL 395

Db 1934 GTKSGTYFVTVSVNNY-----VSDT-----KQVTLIADA---GTAKLASLT 1972

Qy 396 SGKDFALNSQ-----NLVWGEKASLNKLVIATIGEDKVPDGPSTISIKSNH 441

Db 1973 SVTSFVSVTTEGATMTASVTDANGNPVEGIKVNFRTSVTL-----STSVETDDR 2023

Qy 442 GITSVNNYITABAAGEAILTIKVD-----VTQKVKVTTDSRLKLSVKANPKLQVV 496

Db 2024 GFABIL-----VTSEVGLKTVSASLADPKPTEVISRLNLAKADINSATITSLIPEGQWVA 2080

Wed Apr 7 17:36:16 2004

us-09-844-281-1.rsp

```
Qy 497 QNKTLPTFTVTQYDPPFGANTAAIKVLPKTVVAGGLDVVTTDSGSIKTKTIGVTG 556
Db 2081 QD--VAVKAHVNDQGNPILNESVTSAPPEHMTISQ--NIVSTDTHGIAEVMTWTPER 2135
Qy 557 NDVGEGTVHFQNGNGATLGSLYNVNTEGNVAFKNFELVSKVQYQSPDKLDLNVSTTV 616
Db 2136 N--GSYVVKASLANGSSYEKDLVVIDQKLTLSASSELIGVNSPTGATLTATLTSANGTEV 2193
Qy 617 EYQLSKYTSDRVYSDPENLEGEYEVESKNLAVADAKIVGNKVVTGKTPGKVDIHLTKNGA 676
Db 2194 EGQVINF--VTPEGATLSGGKVRTNSSGQAPVLTSNKV---GTYTVTASFH---NGV 2244
Qy 677 TAGKATVEIYQETIAIKSVNFKPVQTFNFKKINIGTV---LELEKSNLDD-----IVKG 729
Db 2245 T-----IQQTIVKVTGNSSTAHVASFIADEPSTIAATNSDLSTLKATVEDGSGNLIIEG 2297
Qy 730 INLTKETCHKVRVVKSGA---EQKLYLDRNG-----DAVFAGDVKLG 770
Db 2298 LTV-----YFALKSGSATLTSLTAVTDQNGIATTSVRGAIKTSVTVSNTTAGGMQIV 2350
Qy 771 DVTV-----SQTSDSALFNFKADLYDTLTTKYTDKGTL-----VFKVLKDKDVITS 816
Db 2351 DITLVAGPADASQSVLKNRS-----SLKGDFTDSAEHLHLVHDISGNPIKVSSEGLEFVQS 2406
Qy 817 BIGSQAVHVNV 827
Db 2407 --GTNAPYVQV 2415
```

Search completed: April 7, 2004, 17:33:32
Job time : 24 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:26:39 ; Search time 24 Seconds
(without alignments)
3338.648 Million cell updates/sec

Title: US-09-844-281-1
Perfect score: 4202
Sequence: 1 AGKSPDPVAGHWAEGSINY.....ITSEIGSQAVHVNVLNPNL 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:**
1: Pir1:**
2: Pir2:**
3: Pir3:**
4: Pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2833.5	67.4	874	2 JC4930	S-layer protein pr
2	719	17.1	814	2 I40048	S-layer protein pr
3	348	8.3	1176	2 A33856	surface-layer 125K
4	265.5	6.3	762	2 A34355	cell surface prote
5	255.5	6.1	1099	2 T14850	S-layer protein pr
6	252.5	6.0	941	2 S29043	cellulase IBC 3.2.
7	242.5	5.8	1036	2 T30311	S-layer protein -
8	240.5	5.7	4919	2 T31105	hypothetical prote
9	239.5	5.7	652	2 B59102	hypothetical prote
10	236	5.6	1109	2 A56143	surface-array prot
11	233.5	5.6	1386	2 AC1533	surface protein (L
12	231.5	5.5	2059	2 D82671	surface protein XF
13	231	5.5	1268	2 B97789	hemagglutinin/hemo
14	231	5.5	1270	2 E85649	hypothetical prote
15	230.5	5.5	2154	2 F83068	hypothetical prote
16	230	5.5	1325	2 A64905	ydek protein - Esc
17	229	5.4	1612	2 AB1347	probable peptidogl
18	226.5	5.4	2020	2 C48339	ABC-type transport
19	225	5.4	1324	2 T18255	endo-1,3(4)-beta-9
20	225	5.4	3029	2 S76109	hypothetical prote
21	224	5.3	1806	2 AF1717	probable peptidogl
22	219	5.2	1116	2 A35129	surface layer prot
23	219	5.2	1343	2 E30893	hypothetical prote
24	218.5	5.2	1175	2 S70843	hemolytic protein
25	217	5.2	1483	2 C97042	probably celluloso
26	217	5.2	1655	2 E37835	hypothetical prote
27	217	5.2	5291	2 F97636	hypothetical prote
28	216.5	5.2	1536	2 A43855	high-molecular-we
29	216	5.1	5188	2 B85547	probable RTX famil

30	215.5	5.1	1635	2 AI0452	hemolysin (importe
31	213.5	5.1	1461	2 E90696	hypothetical prote
32	213	5.1	1343	2 D85724	hypothetical prote
33	213	5.1	1477	2 B43855	high-molecular-we
34	212.5	5.1	1461	2 A85547	hypothetical prote
35	212.5	5.1	1959	2 AG1085	hypothetical prote
36	211	5.0	1084	2 A28555	middle wall protei
37	211	5.0	2256	2 AD1018	large repetitive p
38	211	5.0	3013	2 AB0480	probable invasivi
39	209.5	5.0	2340	2 B71704	cell surface anti
40	209.5	5.0	2554	2 AB3528	extracellular seri
41	209	5.0	404	2 F59097	hypothetical prote
42	209	5.0	694	2 AC2008	hypothetical prote
43	208.5	5.0	5627	2 C83339	hypothetical prote
44	207.5	4.9	1345	2 H90975	hypothetical prote
45	207.5	4.9	1645	2 JN0896	crystalline surfac

ALIGNMENTS

RESULT 1

JC4930
S-layer protein precursor - Bacillus licheniformis
C:Species: Bacillus licheniformis
C>Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 15-Oct-1999
C:Accession: JC4930
Rizhu, X.; McVeigh, R.R.; Malathi, P.; Ghosh, B.K.
Gene:173, 189-194, 1996
A:Title: The complete nucleotide sequence of the Bacillus licheniformis NM105 S-layer-e
A:Reference number: JC4930; MUID:97082965; PMID:8964497
A:Accession: JC4930
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-874 <ZHU>
A:Cross-references: GB:U38842; NID:G1055336; PIDN:AAC44405.1; PID:G1055337
A:Experimental source: strain NM105
C:Comment: This protein is a glycoprotein. It functions as protective coats, molecular
C:Genetics:
A:Gene: olpA
C:Superfamily: S-layer repeat homology
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-874/Product: S-layer protein #status predicted <MAT>
F:157-209/Domain: S-layer repeat homology <SLR3>

Query Match 67.4%; Score 2833.5; DB 2; Length 874;
Best Local Similarity 67.3%; Pred. No. 3.1e-125;
Matches 569; Conservative 98; Mismatches 156; Indels 23; Gaps 7;

QY	1	AGKSPDPVAGHWAEGSINYLDKGAITGKEDGTGPTESIDRASAAVFTKIILNPVDE	60
DB	30	AGKSPDPVAGHWAEDSINYLDKGAIVGKPDGTYGPTESIDRASAAVFTKIILNPVDE	89
QY	61	NAQPSFKDAKNIWSSKYIAAEVAKGVKGDGKGFPEPKIDRASFAFMLVAYNLKDKV	120
DB	90	NAQPSFKDAKNIWSSKYIAAEVAKGVKGDGKGFPEPKIDRASFAFMLVAYNLKDKV	149
QY	121	NGELVTTTDLDDHGWGEKANILNLGI SVGTGGKWEKNSVSRARAAQFTALTDKKYK	180
DB	150	DGTLVTKFDLGHGWGEKANILNLGI SVGTGGKWEKNSVSRARAAQFTALTDKKYK	209
QY	181	KDAQAAYTVDKVSBPTKLTGTGLDKLSADDDVLEGDKVAIAEASTDGTSAVVTLGK	240
DB	210	PENSAKVTNVAAFTQTLTGTGLNKLTAEDVLEGNKAIALBASDKGSAAVTLGK	269
QY	241	VAPNKDLTVKKQGFVTKVFEVKKLAVEKLTFFDDDRAGQAIAFKLNDEKGNADVEYL	300
DB	270	IAPNKELPVKKGNTFIVKYVFEVKKLAVEKLTFFDDDRAGQAIAFKLNDEKGNADIE	329
QY	301	LANHDKVFVANNLDGSPANIPGGEGATSTTKLAVGIKQGDYKVEVQVTKRGGLTVSNTG	360
DB	330	IAGHDVKFVANNLDGTPANIPFEGGTAEETTKLAVGIAEGKYKVEVQVTKRGGITVNTG	389

QY 361 IITVKNLDTPASAIKNNVFPALDADNDGWNVYSGKSGDFALNSQNLVVGKASLNKLVA 420
DB 390 IIEVKNLDAETALKDVPFAVDTKAG-VNYAKFLSGDTFTLNSKTLVAGEKAGIKVVA 448
QY 421 TIAGEDKVPDGSISIKSSNHGIIISVNNYITAAAGATITIKVGDVTKDKVKEKVTDS 480
DB 449 QINKENKVPDPSAISLSKSSNPGVISVKNNGEIKAAAGSATLTVKVGDDVTKTDFVVKTD 508
QY 481 RKLVSVKANPKLQVQNKTLPTVFTVTDQYDGPFGANTAAIKVLPKTVGVAEGGLDYV 540
DB 509 RKLTVKANPDQLKVVDPKELPVTFTVTDQYDGPFGANSAGIKVFPFGVVKV--LDVT 566
QY 541 TTDGSGITGTVIGVTDNDGEGIVHFGNG--GATIGSLVYVWTEGNVAFK--NPELVSK 596
DB 567 TTNEGSGITGSIKVGNGVAGTIHFQNPNASGEGYSLHVTVTKNIGHEAPRLELVSK 626
QY 597 VGOVQGPDPDKLNLNVSTVEYQLSKYTSRVSVDPENLEGYVESKNLAVADAKIVGNK 656
DB 627 AGQGEBAADT--TLGAGTVAYQLSNYTTGTVVADAADLAGYBFRVGNDKIASAKIEGKT 684
QY 657 VVTGKTGPKVDIHLTKNGTAGATVVEIIOETIIAISKVNFKPQVOTENFVEKKINIGTVL 716
DB 685 LKVTGKTAGVTDVILTKDGTAGATITVQENIQTISVKFQVDEVEQFENKRNKIDRVL 744
QY 717 ELEKSNLDDIVKGNLTKETCHKVRVVKSGAEOGKLYLDRNGDAVFNAGDKLGDVTVSQ 776
DB 745 DVKSDKDDVLNGIKLNIHTEHKVRIVDEGTEQGVYLDNRDNATPDGNDVALGYVTAYK 804
QY 777 TSDSA-----LPNF-----KADLYDILTKYTKDGLTVFKVLKDKDVIITSEGSQA 822
DB 805 SNDTVSKEGNDLFRFLTDATNKNDFVKGVTTAFGDKGTIVFKWMDRVAPVTEYGTGA 864
QY 823 VHVNVL 828
DB 865 VTINVI 870

RESULT 2
I40048
S-layer protein precursor - Bacillus anthracis
C:Species: Bacillus anthracis
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jul-2000
C:Accession: I40048; S51695
R:Etienne-Toumelin, I.; Sirdar, J.C.; Dufrot, E.; Mock, M.; Fouet, A.
J. Bacteriol. 177, 614-620, 1995
A:Title: Characterization of the Bacillus anthracis S-layer: cloning and sequencing of
A:Reference number: I40048; MUID:95138020; PMID:7836294
A:Accession: I40048
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-814 <RES>
A:Cross-references: EMBL:Z36946; NID:G1359499; PIDN:CAA85408.1; PID:G634013
C:Genetics:
A:Gene: sap
C:Superfamily: S-layer repeat homology
F:34-88/Domain: S-layer repeat homology <SLR1>
F:95-147/Domain: S-layer repeat homology <SLR2>
F:156-209/Domain: S-layer repeat homology <SLR3>

Query Match 17.1%; Score 719; DB 2; Length 814;
Best Local Similarity 29.5%; Pred. No. 3e-26;
Matches 229; Conservative 126; Mismatches 302; Indels 118; Gaps 27;

QY 1 AGKSPDPVAGHWAEGSINLYDKAITGKPDGTGPTESIDRASAIVFTKILNLPVDE 60
DB 30 AGKTPDPVADHWGIDSINLYVERGAVKNGDKMPFGKELUTRAEATMMQAILNLFIDK 89
QY 61 NAQSPFKAQNTWSSKYIAAEKAGVVKGDGKFNFPYPEGKIDRASFAFMLYSAYNLKDKV 120
DB 90 DAKSPADSGQGYTFPFAAEKAGVVKGTG-NGPEPNGKIDRVSMASLLVEAYKLDTKV 148
QY 121 NGELVTTTPDLDL-HWGEKANILNLGISVGTGKWFENKSVSPAEEAQAQFIATDKYG 179

DB 149 NGTPATKFKOLETLNWSKEKANILVELGISVGTGDQWEPKKTVTKAEEAQFIATDKQFG 208
QY 180 ----KCONAQAYVT---DVKVSEPTKLTLTGTGLDKLSADDVTL-----EGDKAVAIET-- 226
DB 209 TEAAKVESAKAVTTQKVEVKFSK-----AVEKLTKEIDIKVTNKANNDKVLVKEVTL 259
QY 227 STDGTSAVVTLGGKVAFNKDLTKV---KNOSFVTKFVVEVKVLALEKLTFFDDDRAGQA 282
DB 260 SEDKKSATVELYSNLAQKQTVTVVNVKVKTEVAVGSLEAKTIEMADQTVVADP---TA 316
QY 283 IAFKLNDEKGNADVEYLNLANHDKVFNANLDDGSPANIFEGGEATSTTKLAVGKOGDY 342
DB 317 LQFTVVDENGTEV---SPEGIEFV-----TPA-----AEKINAKGEITLAKGTS 358
QY 343 KVEVQVTKRGGLTSYNTGIIITVKNLDTFASAIKNNVFPALDADNDGWNVYSGKSGKDPAL 402
DB 359 TTVKAVYKQKGVVAESKEVKSAGAAVASISNWTVAEQNKAD-----FTSKPFQK 410
QY 403 NSQNLVVGKASLNKLVIATAGEDKVVDPGSGISIKSSNH--GIISVNNYITASAAGEAT 460
DB 411 NNN-VYEGDNAVQ---VELKDQFNATVTKGVEYESLNTVAVVDKATGKVTLSAGKAP 466
QY 461 LTIKVG-----VTKDVKFKV--TTDSKLVSVKANPDKLVQVQNKTLPTVFTVTDQYGD 513
DB 467 VKVTYKDSKGLVSKVTVEIEFAFAQKAMKEIKLRTNVALSTKDVTDLKVAFVLDQYCK 526
QY 514 PFGANTAAIKVLPKGTG-VVAEGGLDVVTTDSSIGTGTIGTGVNDVGEETVHF--QNGN 570
DB 527 BFTAPVTV--KVLDKDGKELAEQKLEAKVYNKELV---LNAAGQEAAGNYTVLTAKSGE 580
QY 571 GATLGSLVNVTEGNVAFNPFELSVKVGQYQSPDKLNLNVSTVEYQLSKYTSRVS 630
DB 581 KEAKATLALEL-KAPGAFSKFEV-----RGLEKELDKYVTEENQK 619
QY 631 DPENLEGYVESKNLAVADAKIVGNKVVVTKGTPGKVD-----IHLTKNGATA 678
DB 620 NANTVSVLPVDANGLVLKGAESAELKVTITNKEGKEVDATDAQTVQNNSVITVQCGAKA 679
QY 679 GKA-TVEIVQETIATKSNVFPVQVOTENFVEKKINIGTVLEKSNLDDIVKGINL 732
DB 680 GETYKVVVLDGLITTHSFVKVDTAPTAK-----GLAVEFTSTSKVAPNADL 729

RESULT 3
A33856
surface-layer 125K protein precursor - Bacillus sphaericus
C:Species: Bacillus sphaericus
C>Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 15-Oct-1999
C:Accession: A33856
R: Bowditch, R.D.; Baumann, P.; Yousten, A.A.
J. Bacteriol. 171, 4178-4188, 1989
A:Title: Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer pro
A:Reference number: A33856; MUID:89327128; PMID:2666389
A:Accession: A33856
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1176 <BOW>
A:Cross-references: GB:M28361; NID:G341911; PIDN:AAA50256.1; PID:G556012
C:Superfamily: S-layer repeat homology
F:93-147/Domain: S-layer repeat homology <SLR2>

Query Match 8.3%; Score 348; DB 2; Length 1176;
Best Local Similarity 24.2%; Pred. No. 1.2e-08;
Matches 236; Conservative 123; Mismatches 307; Indels 310; Gaps 55;

QY 13 WAEGSINLYVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDENAQPSFKDAK-N 71
DB 42 YAKEAVQSLVDAGVIOGDANGFNFLKTSRAEATIFTNALEAEAGDV--NFKDVKAD 99
QY 72 INSSKYIAAEKAGVVKGDGKFNFPYPEGKIDRASFAFMLYSAYNLKDKVNGELVTFED- 130
DB 100 AWYDAIATAATVNGIPEGVSATFAPNPKLTREAAKILVDFAFELEGE--GDL-SBFADA 156

T14850

S-layer protein precursor - Bacillus stearothermophilus
 C:Species: Bacillus stearothermophilus
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14850
 R:Jarosch, M.; Egelseer, E.M.; Mattanovich, D.; Sleytr, U.B.; Sara, M.
 submitted to the EMBL Data Library, April 1999
 A:Description: Nucleotide sequence of the coding region of *sbsC*, the S-layer gene from *B. stearothermophilus*
 A:Reference number: Z18239
 A:Accession: T14850
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1099 <JAR>
 A:Cross-references: EMBL:AF055578; NID:G4581955; PID:G3025826; PIDN:AAC12757.1
 A:Experimental source: ATCC 12980
 C:Genetics:
 A:Gene: *sbsC*

Query Match 6.1%; Score 255.5; DB 2; Length 1099;
 Best Local Similarity 22.0%; Pred. No. 0.00023;
 Matches 214; Conservative 125; Mismatches 330; Indels 303; Gaps 49;
 QY 58 VDENAQPSKDAKNWS-----SKYIAVEKAGVVGKDG-----KEN 94
 Db 56 VTEGQ--FPDKQVYAYNAKQAYANAVAVNKAAGAKKADAYLADLQIYETVFKAN 113
 QY 95 FYPEG-----KIDRASFASML-----VSAYNLK----- 117
 Db 114 --PKSGEARVATYIDAYNATKDRQELKAAVDADKKAEBLYHKISVELKTRTVIL 171
 QY 118 DKVNGE-----LVTTTF-----EDLDHNGEKKANIL-----IN 145
 Db 172 DRVYQGSRELLRSIFKADAQALRDLIYDITVAMKAREADAVKAGNLKAKAALDOVN 231
 QY 146 LGISVGTGG-KWEPKNSVSRBAQFIATLDPKKGKDNQAAQYVTDVVKVSEPTKLTGT 204
 Db 232 QYVSKVTDAPKAELOKAAQDAKAAVEAALTPK-----VESVAIDSTSFVKVTF 280
 QY 205 -GLDKLSA-----DDVTLEGDKA-----VAIEASTDGTSAVTLGGKVAENKDLTVKVN-- 253
 Db 281 KPVDKATAPKPFSTILKGTETKLPKSVSEBSGLTATVTLTYLTVDTGKTYIVTVSGLK 340
 QY 254 -----QSFTVKFYVEVKLAVEKLTDPDDRAQAGIAFLKNDSEKNADVEYLNLANHDV 306
 Db 341 DTAGKEFETSTNEFTY--NKPVPASITFNFKLPEDSAVDL-----TKYVT-----V 385
 QY 307 KPVANN-----LDGSPANIFEGGEATSTTGKLAAGIKQGDYKVEYQVTKRGSLTVSNT 359
 Db 386 KDAAGNVKSGELEFTSEKLTQKFPINTGKS-----VIVNAIVG--TNVIT 434
 QY 360 GIITVKNLTPASAIKNVVFALDADNDGVNYY-----GSKLSGKOPALNSQNLVVGEXAS-- 414
 Db 435 GNVILAVEDEKAAEVSEL--KLTKDNKEVTVLYANGNAFDQGNQISSGTLTLTAKPKDQ 492
 QY 415 -LNKLVAITAGDKVVDPGSISIKSNHGIISVY--NNYITAEAAAGEATLTKVGDVTKDV 472
 Db 493 YGNELTGKVGTD-----YTFESLNEVLAVAPDGSVPIVPGTALVKVKYGEVTKTI 545
 QY 473 KFKVTDSDKLVSVKAMP-----DKLQVQNTKLPVTFVTT--DOYGPDPGANTAAIK 523
 Db 546 P-----VTVKANPVLFTIADVSTGVSVAKGQKATPKVTLKQYGNKFTGNVNTS 595
 QY 524 EYLPKTVVAEGGLDVTWTDGSGIGTKTIGVTGNDVGEVTFVFGNGGATLGLSYNVTE 583
 Db 596 D---KTEIA-----TVSVNSGIGQSEYTVTVNGVAEGSTTIKISGTEKVPVNVVA 646
 QY 584 GNVAPKNFELSVKQYQSGPDKLDLNVSTTVETVYQLSKYTSDRVYDPPENLGEYVESK 643
 Db 647 GG-PVANYQI--KV-----LDGKID-----KSGATESPANNDVQLKVIYAVDAN 686
 QY 644 NLAVADAKIVGNKVVVTKTP-----GKVDIHLTKGA--TAGKATVETVQ 687
 Db 687 GNIVGD---ITNDVTITSEATDTNGVIVNASKSTANGDTVVVITDNGSKVGTETLVKL 743

QY 688 ETIAIKSVNFKPVOT---ENFVEKKINI-----GTVLELEKSNLDDIVKGIN----- 731
 Db 744 GTVTLGTVEVIDTTLKATVTKADLIELDAADNGDALAKLANLD--IKQNGNPMV 801
 QY 732 --LTKETQHKVRVK-----SGAEQGLYLDKNGDAVFNAG----- 765
 Db 802 DSAATPNTNEKLQAKSVLGSIVSDTSVTSVGSNVNKLDDASISGLAVKKAGTTLTL 861
 QY 766 ----DVKLGDVTVSQTSDSALPN-----FKADLYDTLT---TKY---TDK---GTLVFKVL 808
 Db 862 VNEEDSKIAPAIATVKAAPATQGVTVTGLDLPVGVTKTFTATDKIKSGHKLIVAV 921
 QY 809 KDKDVIITSEIGS 820
 Db 922 DDSAVPAPAVGT 933

RESULT 6
 S29043
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: Bacillus sp.
 C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
 C:Accession: S29043; PC4404
 R:Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
 J. Gen. Microbiol. 136, 1327-1334, 1990
 A:Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from *Bacillus* sp.
 A:Reference number: S29043; MUID:91037937; PMID:2230718
 A:Accession: S29043
 A:Molecule type: DNA
 A:Residues: 1-941 <OZA>
 A:Cross-references: EMBL:M27420; NID:G142664; PIDN:AAA22304.1; PID:G142665
 R:Shirai, T.; Yamane, T.; Hidaka, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozaki, K.; Ito, S.; Biochem. 122, 683-685, 1997
 A:Title: Crystallization and preliminary X-ray analysis of a truncated family A alkaline cellulase from *Bacillus* sp.
 A:Reference number: PC4404; MUID:98060488; PMID:9399567
 A:Accession: PC4404
 A:Molecule type: protein
 A:Residues: 228-584 <SHI>
 A:Experimental source: strain KSM-635
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
 A:Pathway: cellulose degradation
 C:Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homology; Thermotoga
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:41-95/Domain: S-layer repeat homology <SLR1>
 F:101-153/Domain: S-layer repeat homology <SLR2>
 F:164-219/Domain: S-layer repeat homology <SLR3>
 F:766-908/Domain: Thermotoga xylanase A amino-terminal repeat homology <TYA>

Query Match 6.0%; Score 252.5; DB 2; Length 941;
 Best Local Similarity 21.3%; Pred. No. 0.00025;
 Matches 193; Conservative 109; Mismatches 291; Indels 311; Gaps 42;
 QY 5 FPDYPAGHWAEGSNLYLVKGAITGKPDGTYGTESIDRASAAVITKILNLPVDENAO 64
 Db 41 FSDVKYTSWSPFYKIDLYEGEIVTIGTSATTSFSDSVTRAQFTVMTLRLGLGEASSKDP 100
 QY 65 SFDAKNIWSKIIAAVEKAGVVGKDGKFNPEYEGKIDIRASFASMLVSYNLT----- 116
 Db 101 -FKDRKN-WAYKEIQAAEYAGIVTGKTNGEFAPENITREQMAAMAVRAYEYLENLSLP 158
 QY 117 ---KDKYNGELVTFP--EDLDHNGEKKANILINLIGISVGTGGKWEKNSVSRRAAAQFIA 172
 Db 159 EEQREYNDSSISTPQADAV-----QKAYVLELM--EGWTDGVFQPKRSTRQSAKVIS 211
 QY 173 LTDKYYGKQNAQAYTVDKVSEPTK-----LTLTGTGLDKLSADDDVTLLEGKAVAIK- 225
 Db 212 TLLKWKVASHD--VLYHTEA-VKSPSEAGALQLVELNG-----QLTLAGEGDTFVQJ 259
 QY 226 --ASTDGT---SAVYTLGGKVAENK-----LTVKVNQSFVTK-----FVYEVKILA 268

Db 260 RGMSTHGLQFGEIIVNENAFVALSNDMGSNNIRLAMYIGENYATNPEVVDLVEGIELA 319
 QY 269 VEKLF-----DDDRAG-----QATAPKLNDKGNADVEYL--NLANDVK 307
 Db 320 FEHDMYVIVDWHVAFGPRADVSGAYDPFEEADHYKHDPKX--HYIWEIANEP-- 374
 QY 308 FVANNLDOSP--ANIFEGGEATSTTKLAVGI--KQGDYKVEV---QVTRKGLTIVSNTG 360
 Db 375 --SPNNGGPGLTNDKGEWAKEYAEPVEMLEKRGDNMILVGNPNWSRQPLDSADNP- 431
 QY 361 IITVQNL-----DTPAS-----AIKNVFPALD----- 382
 Db 432 -IDAENIYSVHYFTGSHGASHI GYPECTPSSERSNWMANVYALDNGVAVFATWGTQ 490
 QY 383 ANDNG-----VNYGSK--LSGKDFALNSONLVV--EKASLNKLVAIAGSDKV 429
 Db 491 ANGDDGPGYFDEADVWLNFLNKHNTSWANWSLTNKEISGAFTPELGRDTAT-----DL 544
 QY 430 DPGSISIKSNHGIIISVNNVITAEAGEATLTIKVGEDTKDVK-----PKVT 477
 Db 545 DFGANQVAPEE--LSLSGEVVRARIKIEVTPI---DRKFTKLWDFNDGTTQGFQVN 599
 QY 478 TDS--RKLVSKANPDKLOV---VQN----- 498
 Db 600 GDSPNKESITLSNNNDALQIEGLVNSNDISEGNTWMDNVR--SADGWSNVDILGATELTID 659
 QY 499 -----KTLPTVFTVTOYGDPPFGAATAA--KEVLPKTVGV 532
 Db 660 VIVEPTTVSIAAIPQGAAGWNPTRAIKVTEDDPFSGDGYKALVITSEDSPLI 719
 QY 533 AEGGLD-----VTTDSGSGTGTIGVGNVNDGEGTVHQNGGATLSLY----- 578
 Db 720 ATSPEDNTWSNIIIFVGTEADVISLNDNITVSGTEIEBIEVHDEKGT--ATLPSTFDGTR 778
 QY 579 -----VNVTEGN-----VAFKNFELSVKVGQ 599
 Db 779 QGMDWHTSGVKLTALTEFANGSNALSWEYAYPEVSPDGCWATAPRLDFWKDELVRGTS 838
 QY 600 Y-----GQSPDTKLDLNVSTTVYQLSKYTSRVSYPDENLEG 637
 Db 839 YISDFYIDAVRASEGAISINAVFPFANGYQVETPTFIDLTELDASATVTS--ELVH 896
 QY 638 YEVE 641
 Db 897 YEVE 900

RESULT 7

T30311
 S-layer protein - Clostridium thermocellum
 C:Species: Clostridium thermocellum
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30311
 R:Lemaire, M.; Miras, J.; Gounon, P.; Beguin, P.
 Microbiology 144, 211-217, 1998
 A:Title: Identification of a region responsible for binding to the cell wall within the
 A:Reference number: Z20818; MUID:98129094; PMID:9467913
 A:Accession: T30311
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1036 <LEM>
 A:Cross-references: EMBL:U79117; NID:g3493463; PID:g3493464; PIDN:AAC33404.1
 C:Genetics:
 A:Gene: slpA

Query Match 5.8%; Score 242.5; DB 2; Length 1036;
 Best Local Similarity 19.4%; Pred. No. 0.00086;
 Matches 206; Conservative 152; Mismatches 351; Indels 353; Gaps 52;

QY 8 VPA-----CHWAGSGINYLVDKALTKPDTGYCT--ESIDRASAAVITKI----- 53
 Db 22 VPAPADSPSYKEAEILVRL--GLYKGTSEYTVNLEGKLDROTGVVMLLRFLFGQEDDA 79

QY 54 LNLVNDENAO---PSFKDAKNI--WSSKYIA-AVERAGVGVKGDKENFYFEGKIDRASFA 107
 Db 80 LEIPWDEAAQTAAKFKDAADIAADWQROQVAYAVEK-GYVKGYPDGTFLENADLNGLAFC 138
 QY 108 SMLV-----SAYNLKDKUNGELV----- 125
 Db 139 SLIIQQIAGYDGFVDFDAAYKLOEFGLTAEQABAFNNKNGINRDSMWGIAFSALQAVTK 198
 QY 126 ---TTFEDLLDHWGEEKANILINLG-----ISVGTGKWEF----- 158
 Db 199 ATGKTVIEVLVEN-GNYSKELALIEGLVLLKAIKEVKALDAVKVOVGK---EVLPEEVE 253
 QY 159 ---NKSVSRAEAQFIALTDKYKGD-----NAQAY-----VTD 190
 Db 254 VYEDDTTEKLAWEVFPVDTSEVGEQIEGTIKAGSLAYREP KATLVIVTPEEQVVD 313
 QY 191 VKSEPTKLTITGTDKLSADD---VTLEGDKAVAEIASTDGTSAVTLGGKVAENKOL 247
 Db 314 VKAPNLKEIIVIEFNGEVASKADEKSSYSVEDNTIELVTVSEDKTTVTTLTVAGANTAEIEI 373
 QY 248 TVKVNQSFVTKFVVEYVKLAVEKLTFFDDDRAGQATAFKLNDKGNADVEYLNLA---NH 304
 Db 374 EVTKT---ATGLKEEVTKTWPA-----DYENPEAESIALIGPNSF 412
 QY 305 DVKEFVANNLDGSPANIPEGG-----EATSTTGKLV--GIKQGDYKVEVQVTK- 350
 Db 413 EIKFSEPVQSSDAEVLVNDGTVYVSEKLSQDYRTLTVELGVSLNEGTYKVKVGYRD 472
 QY 351 -RGLTIVSNT-GIITVKNLOTASAKN-----VVEALDADNDGVVNVGSKLSKGDFA 401
 Db 473 YAGNIMRTKTFDLEKYKDDTPPTAKVKEATONKVVIFNEPATRDG-----YSDEAA 525
 QY 402 LNSQ-----NLVVGEKASLNKLVAIAGEDK-----VVPDGSISIKSNH 441
 Db 526 LTRDYFYQTVSSMKPTKVASDN---NKVYTLFSESDQNDGGYFVYLLPVGNVTIT--- 578
 QY 442 GIISVANNYITAAAG-----EATLTIKVGQVTKDVKFKVTTDSRKLVSVKANPKLQ 494
 Db 579 -LLEVEDDDDAVDWANGKLESCLKLTATVAADNEATVK-SVTAAE- -DKIV 627
 QY 495 VVQNTKLPVTFVTTDQYDPPFGAATAIKE--VLPKTVVVAEGGLDVVTTDSGSI----- 547
 Db 628 VV-----PSEDVNEQAKDNVYVKKDGKEIDTAISSITVDSNETKVTIV 673
 QY 548 -----GTKTIGVGNVGEQVHFO----- 567
 Db 674 LDEKLSGKVTIDIKIGKOTSVSENEKAVTIEFVTDKTAPIEFTVFDNVIYRYSE 733
 QY 568 ---NGNGATLGLYNNVTEGN---VAFKNFELV-SKVQYQGSPTKLDLNVSTTVYEQ 619
 Db 734 AMSTKNGSVLNKDNKYLVDNDKKVEIKKIELFGSDKNKVRITVDSVDLNL---VDYE 789
 QY 620 LSKYTSRVSYPDEN-LEGYEVESKLVADAKIVGNKVVTGKTPGVDIHLTKNGATA 678
 Db 790 L---TIANVEDAGNAISADVAKKLSBEQAEV-SEIRISKT--EIEIVINK---IL 840
 QY 679 GKATVEIVQETIAIKSVNFKPVQTEVFKEKINIGTVLEKSNLDD---IVKGINLTKE 735
 Db 841 DKATVE-----KTDFEVERGSNKVALTRISSITYDDGKTIKGV----- 879
 QY 736 TQHKVRVVKSGAEGKGLYLRDNGDAVFNAGDVK-----LGDVTVTSQTSLSALPNFKADL 789
 Db 880 LPDAVRPANSQ-----DITGYTLIVGEIKSDTGKEMATGAVSKFVDD---KPAF 928
 QY 790 YDTLTTKYTKGTGLVFKVKDKOV--ITSEIGSQAVHNVNLN 829
 Db 929 VSVANGVYGDASKKGFTLTDFEDIKFLNNSAGLGATDLVIKN 970

RESULT 8

T31105
 hypothetical protein 2 - Haemophilus ducreyi
 C:Species: Haemophilus ducreyi

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T31105
 R: Ward, C.K.; Lambley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
 J. Bacteriol. 180, 6013-6022, 1998
 A: Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
 A: Reference number: 220984; MUID: 99030326; PMID: 9811662
 C:Accession: T31105
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-4919 <var>
 A: Cross-references: EMBL: AF057696; NID: g39290021; PID: g3929023; PIDN: AAC79761.1
 C: Genetics: 18pA2

Query Match 5.7%; Score 240.5; DB 2; Length 4919;
 Best Local Similarity 20.4%; Pred. No. 0.0099;
 Matches 206; Conservative 147; Mismatches 353; Indels 305; Gaps 51;

QY 13 WAEGSINY-----LYDKGAIKCP--DGYGPTSIDRASAIVFTKILNLPVDENAQPS 65
 DB 294 FAAGNUTYDVRDVRNRPKPTIDTRKDNIAISGESAGSAGMGRNKFIVTDKG--- 350
 QY 66 FKDAKNIWSKYIAAVEKAGVKGDKENFYPE---GKIDRASFAFMLVSAYNLKDVKNG 122
 DB 351 -----AGVNHQGVIFAEDDINILTDGNSRLNKVYADYVRVVGKQIELANNG 397
 QY 123 ELVTTFEDLLDHGKEKAN-----ILNLGISVGTGKWEKPKSVSRABAAQFIALTCKY 178
 DB 398 QIHADQQLINATGHVKLNDGSSVLSNNLGLISALNLTLENATVS-ANNLSFRVNTNKL 456
 QY 179 GKDKNAQAVY-----TDVKVSEPTKLTITGTGLDK----- 208
 DB 457 NNLKSVSARAADLQSGNLNLDKASVLAKHLTLNISDVLNNQSKLSANNLKIKKVRDLN 516
 QY 209 -----LSADDVTLLEGKVAIAEASDGTSAVVTLG--GKVAIPKDLTVKVKQSF-VTKF 260
 DB 517 LNNSELSANNLTLNNTSNWTLKQKFTAGNLTNTNVTNLTNDSELAANNLTNTNKN 576
 QY 261 V--YEYKLAWEKLTPE--DD-----RAGQAIAPKLNDEKGNADVEYLNLANHDVK 307
 DB 577 VTLNDASKLSANKLDLNTDNTVNLNSKSTLSAGE-LTEK-----KVKNVTL-NNDSE 626
 QY 308 FVANNLDGSPANIFEGGEATSTTKLAVGIGKGVQVQVTKRGGLTVSNTGIITVKNL 367
 DB 627 LAANNLS-----LNASHNVTLNKSKLSAQAKADIKAV-----NLTLNITTELTAKNL 673
 QY 368 DTPASAIKXVVFALDADNDGVN-----YGSKLSGKDPAL-----NSQNLVY----- 409
 DB 674 DINSTTI-----TNGGTIAGIFANITTEKLNKKEKALILAPQNLNFTVNGSHYENK 724
 QY 410 GEKASLNKLVAITAGEDKVVDPGSIKSSNHGIIISVNNY-----ITAEAA 456
 DB 725 GDIVSKDKATVTSKNSDFTSNGSKLVNAQNLKVN-VNNFTISQGGDITILGNVTLNAS 783
 QY 457 GEAT-----LTIKVGPD-----VTKDV-----KFKVTTDSKLVSVK--ANPD 491
 DB 784 GTFTNSGNLTVTATLDVGDIQNTKNGLTGVEDLHIKSKTKITNDG-KLISIKNLNLS 842
 QY 492 KLVQVQNKTL-----PVTFTVTDQYDPPFGANTAAIKVLPKTV-----VAEG----- 535
 DB 843 EADFINNGTLGGLGALKATKGNF-----TNKEKAILASNLSDISVABGKTPFNNGT 895
 QY 536 -----GLDVVTVDG-----S-GTKTIGVTGNDVGEQTV-----HFQNG 569
 DB 896 IESGKNLNTTGAFLNVDNATIRSPGLNLTSTGNSVNGTLISNERLNTISAANFTNE 955
 QY 570 NGATL--GSLVYNTVEGVNAPKPFELVSKVQY--GOSPDTKLDVAVSTTVEY----- 618
 DB 956 SNGTVSNGLLNIIAQGNITKN--LIASRQQLNLTAVADNITNDSNISKNIAVLHSLG 1013
 QY 619 QLSKYTSRYSPPENLEGYVESK-NLAVAQAKIVGNKVVTGKTPGKVDIHLTKNGAT 677
 DB 1014 NISLNSKDVYNLGEIYAGNINISVKAHQKNDVKLMGD---ITTKT-----K 1057

QY 678 AGKATVEIVQETIAIKSVNFKPVQTFENFVKKINI-GTVLELEKSNLDDIVKGINLTKET 736
 DB 1058 EQASIKYLYQ---ASNGHGFNDGSSGYSEGDINIKGKPADLD-----NKL 1101
 QY 737 QHKVRVVKSGAEOQKLYLDNRGDAVFNAGD-----VKLGDTVTV---SQTSDSALPN- 784
 DB 1102 VQRI-----GKIYAGR--DLTFNKSAGGKSEIINRGITINVKVKKLSYSDVSFPENN 1150
 QY 785 ---PKADLYDLTTLTKYTDKGLFVKVLKDKDVTITSEIGSOAVHVNVLNPN 832
 DB 1151 MQSQKVDLY-----TKIFEAKSDIELFTKNTGTHPVLNPKSNNN 1190

RESULT 9

B59102
 Hypothetical protein pX01-90 - Bacillus anthracis virulence plasmid pX01
 C: Species: Bacillus anthracis
 C: Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
 C: Accession: B59102
 J. Bacteriol. 181, 6509-6515, 1999
 A: Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbor
 A: Reference number: A59091; MUID: 99445483; PMID: 10515943
 C: Accession: B59102
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-652 <var>
 A: Cross-references: GB: AF065404; NID: 94894216; PIDN: AAD32394.1; PID: G4894306
 A: Experimental source: Strain Sterne
 A: Note: similar to hypothetical protein; PFB0765w (980 aa); Plasmodium falciparum (AE00
 C: Genetics:
 A: Gene: pX01-90
 A: Genome: plasmid

Query Match 5.7%; Score 239.5; DB 2; Length 652;
 Best Local Similarity 23.8%; Pred. No. 0.00063;

Matches 114; Conservative 80; Mismatches 186; Indels 97; Gaps 19;

QY 4 SPFDYPAGHWAEGSINYLVLDKGAITGKPDGYGPTSIDRAS-AAVTF-TKILNLPVDEN 61
 DB 44 AFLDVPESHANVAIINLLKKEIIVGYGNKFGGLGDSVTREQVAIVYRIKFPKEVEAE 103
 QY 62 AOPSFKDAKNWSSKY---IAAVEKAGVVGDKGENFYPEGKIDRASFAFMLVSAYNLKD 118
 DB 104 AONPRDV-NKSSNIFLDEILTLTKMGIFSGDEKGNFRPKSPISRAEMTVIKNAFNI-- 160
 QY 119 KYNGELVTTTFEDLL-DHWGEEKANILINLIGISVGTG-GKMEPKNSVSRABAAQFI----- 171
 DB 161 PVVGK--HNFRDVAKYWANDAISALQSNQLVSGTGNGLPEPKSVTVREYAQFI FNVLN 218
 QY 172 ---ALTDKKYKKNQAAQAVYTDVKVSEPTKLTITGTGLDKLSADDVTLE-----GKAV 222
 DB 219 FDELVNKQKLDLNNKRDVINQKIDE-----FDKLSQKDLERMLELNQKLS 268
 QY 223 AIEASTDGTSAVVTLLGGKVPANPKOLTVKV-KXQSFVTKFVYEVKGLAVEKLTFFDDDRAGQ 281
 DB 269 QLKQO---SPQLQDLKXKLKESQSRLLELNKKDNRLELNSEIKKL-----NDRKAE 317
 QY 282 ATAF-----KLNDKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATS 328
 DB 318 LLSLIMELIKQOSEPDKIKNEKODLNKKREDLIN-----RIASEKELAK 362
 QY 329 TTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTGII-----TVKNLDTPASAIKN--- 376
 DB 363 KKAELNLTKEV-LFKVQALNKKSQYLYIYNKLDNELRELADRYKNSDKNLSLKNHIG 421
 QY 377 -----VVPALDADNDGVNYSKLSGDKDFALNSQNLVVGKASLNKLVAITAGEDK 427
 DB 422 EYNKQLEKIENELEECNKIKDNTKQALAEFPKSNKKQOESELSVLQNLKIDELGKRHK 480

RESULT 10

```

Query Match      5.6%; Score 233.5; DB 2; Length 1386;
Best Local Similarity 21.9%; Pred. No.0.034;
Matches 197; Conservative 121; Mismatches 339; Indels 243; Gaps 47;

QY   27 ITGKPD---CTYGPTESIDRAAAVIFTKILNLPDEN-----AQPSEKDAKWISSK- 76
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   556 VTGEVDVNTPDGYEIMYTIDGVSTKIIVT-----VKEDKSIEAKOSILIYGTWNSKD 609
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY   77 YIAAVEKAGVVGKGKENFYPEKIDRASPASMLVSAYNLKD-KVNGELVTT--FEDL 131
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   610 NFISATDKG-----NPVDFDKIVEGTVMVTIKPGTNKV 643
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY   132 LDHNGEEKANILLIGISVGTGKEWPKNVSRAAQFIALTDDKYCKKDNAAQYTDV 191
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   644 TVLYGNOSKEVTITV-----KADQSTLEAKOSIITYGDKNWAKDN----- 683
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY   192 KVSEPTKLTITGTGLDKLSAD--DVTLGEDKAVA;EASTDGTSAVVTLGGKVPANKDLTV 249
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   684 -----FITADKDGNPVDFDKDIEVEG-----TVDTTKPGINKTIYYIGNL--SKEVTV 729
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY   250 KVK-NQSPVTKFVYEYVKKLAV-----EKLTFFDDRAGAIAFKLNDEBKNAD---- 295
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   730 TVKANQA-----TLEAKDSALYVGDGTWNSKDNFISATDKDGTVPDFDKDIVEGTVDTTKA 784
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY   296 ---VEYL--NLANH---DYKFVANLLDGS PANIPEGGE-----ATSTTCCLKAVGI 337
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   785 GTNKVTHLYGNOSKEVTNVKADQTTFLEKOSVIYTGDKWAEDNFISATDKGN-SVGF 843
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY   338 KGQGYKVE--VQVTKRGGLTVSNSTGITVKNLDTPASAIRKNVPALDAD-----NDGW 389
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   844 K-DIKVEGTYDITTKAG-----TNKVTYVYANOSKEVTITVKAQDAOTLEAKOSII 891
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:06 11 November 2014

QY 620 LSKYTSRVSYPENLEGVESEKSLAVADAKIVGNKVVVTGKTPGKVDHLT-KNGATA 678
 Db 713 SSGTMD-----NNATGFASDKNLSLE-----VNSLTNYGWSGKGDVTVNNGNLY 762
 QY 679 GQATV--EIVQETIAKSV-NFKPV-----QTEPVKKINIGTVLEL 718
 Db 763 NNTAAEKGDLAALNGIENKWDISAGGDTMTNTRHVTNNSNMVGNIVINAVNDI 822
 QY 719 EKSNDLDDIVKGINLTKETQHKVRVVVSGAEQGLYLDNRNGDAVENAGDVKLGDTVTSQTS 778
 Db 823 --NNGRNVSDADL-----NVTKGNLYNLYVMVGYGDIALSANSVANNAATTIATG 872
 QY 779 DSALP-----NFKADLY---DTLTTK-----YTPKG-----TLVFKVLKD--KDVIT 815
 Db 873 DIIIDSKGNVGNRGNLHALNGVLSVGNKNNLNDNGEIRGYGDVTLTALTONYDSYKSLT 932
 QY 816 SEIGSQAVHVNVLNN 830
 Db 933 SETGDVTLTANIVDN 947

RESULT 15
 F83068
 Hypothetical protein PA4625 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83068
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Harbig, K.; Lim,
 ;; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: F83068
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2154 <STO>
 A:Cross-references: GB:AE004876; GB:AE004091; NID:g9950869; PIDN:AAG08013.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4625

Query Match 5.5%; Score 230.5; DB 2; Length 2154;
 Best Local Similarity 22.3%; Pred. No. 0.0089;
 Matches 216; Conservative 112; Mismatches 327; Indels 313; Gaps 49;

QY 43 RASAAVIFTKIINLPVDENAPQSPFKDAKNVWSSKYIAAVEKAGVYVGDKGKGFYPEGKID 102
 Db 864 QATASTATQASN--VGQYAITGNANGSEYFSQRY-----QLVRQDGLTVTP----- 909
 QY 103 RASFASMLYSAYNLKDKVNGELVTTPE-----DILLDHWEKEMANILINLGI 148
 Db 910 ----AQLIYSA--DAKTKVYGDADPTLTQVSLKNSDPAAGVLSGNLGRVAGENVGYI 964
 QY 149 SVGTGKWEPNKSVS-----RAEAAQIALTDKK---YKKQNAQAY-VTDVKVSEPTK 198
 Db 965 LQGGGLNLTANTYLSVYNGDLRITPAQLNVIADAKTKVYGLDLPALTYQVSLGRGDTAG 1024
 QY 199 LTLTGGLDKLSADVDTLEGDKAVAEASTDGTSAVTLGGKVPNKDLTVKVK-NQSFV 257
 Db 1025 AVLNGSLSRVAGENVGYG-----INQSGGLVSSNYTLNYGNNLTI 1068
 QY 258 TAPVVEYKLAKEKLTFFDDRA--QQAATFKLNDEK-----NADVEYLN- 300
 Db 1069 TRALLNVIADAKTKVYGDADPALTYQVSLKNGDTAGAVLNGGSLSRVAGENVGYGINQ 1128
 QY 301 -----LANHDVKFVANNLDSGPANIFEGGEATSTTGKLAIGIKQGDYKVEVQVTKRGGL 354
 Db 1129 GGLGILLSANYDLSYQGNL-----TITKALLNV---IADAKTKVYGDADPSL 1172
 QY 355 TVSNIGIITVKDLTPAAGAKNVVFPALDADNDGVNVYVSGSKLSGKDFALNSQNLVV---GE 411

Search completed: April 7, 2004, 17:33:00
 Job time : 29 secs

Db 1173 TYQVSGL---KNGDTAGSILTTGGLNRAAGENVGV--YG--INQGD LALNSGNVYDLSYQGN 1225
 QY 412 KASLNKLVATTIAGBDKV-----VDP-----GSI 434
 Db 1226 NLTTIKALLNVIADAKTKVYGDADPSLTQVSGLKNGDTAGAVLNGGGLVRSVGENVGY 1285
 QY 435 SIKSNHGIIISVNNYITAAAGAEALTIKVGDTVTKDKVKVTTDSRKLVSVKANPKLQ 494
 Db 1286 AIOQGGGLVLS--GNYDLAYQGNLTIITKALLNVIADAKTKVYGD-----ADPS--- 1332
 QY 495 VVQNKLTLPVTF-VTTDOYQDPFGPANTAAIKEVLPKTVVAEGGLDVVTTDSGSIGTKTIG 553
 Db 1333 -----LITQVSGLKNGD-----SAGSILTTGGLNRAAGEN-----VG 1363
 QY 554 VTGNDVSEGTVHFON-----GNGATLGSLYVNVVTEGNVAFKPFELVSKVGYQSQPDTK 607
 Db 1364 VYGINQGD LALNSGNVYDLSYQGNLTIITKALLNVIAD-----AKTKVYG----- 1407
 QY 608 LDLNVSTTVVEYQLSKY-----TSDRVYS-----DPENLEGYEVESKNLAVA-----DAKI 652
 Db 1408 -DADPSLT--YQVSGLKNGDTAGAVLNGGGLVRSVGENVGNVYAIQOGLGLVSGNYDLAY 1464
 QY 653 VGNKVVT-----GKTPGKVDIHLT-----KNGATAGKATVEIVQETIAKSV 695
 Db 1465 QGNLTIITKALLNVIADAKTKVYGDADPSLTQVSGLKNGDTAG-----AVLNGG 1514
 QY 696 NFKPVQTEFVEKKINIGTVLELEKSNLDDIVKGINLTRETQHKVRVYVKSAGAQGLYLD 755
 Db 1515 SLRSVAGENVGYGINQGD-LALNSGNVYDLSYQGNLTIITKALLNVIADA--KTKVYGD 1570
 QY 756 -----RNGD--AVFNAGDV-----KLGDVTYSQ----- 776
 Db 1571 ADPSLTQVSGLKNGDTAGAVLNGGGLVRSVGENVGNVYAIQOGLGLVSGNYDLAYQGN 1630
 QY 777 -TSDSALPNFKADLYDTLTTKYTDKG-TLVFKV--LKDKD-----VITSEIGSOA---VHV 825
 Db 1631 LTIITKALLNVIADAKTKVYGDADPSLTQVSGLKNGDSAGSILTTGGLNRAAGENVGV 1687
 QY 826 NVLNNPML 833
 Db 1688 YGINQGD L 1695

QY 121 NGELVTTFTEDLLDHGEEKANILNLGIVSGTGKWEPEKNSVSPAAQAQIALTDKYYK 130
 Db 150 SGELEINFSDLKHGCKDANILVALGINTGNGWEPKSVTRAERAKFIATKDMQFGQ 209
 QY 181 KONAQAYVTVKVSPTKTLT- GTGL-DKLSADVTILEGDKAV---AIEASTDGTSAV 235
 Db 210 K-AEAKVESIKIEINAKIEIKVTEGVDVTAANPAFVVEGSKELDIEKVELSKDKKSATI 267
 QY 236 TLGGKVPAPKDLTVKQNSPVTKFVYEVKKL-----AVEKLTEDDDRAGQAIAPKLN 288
 Db 268 TLKNAL-----VNKAQYVA-HVKDVKSVDGKDPKALEVIFFFDEVAPTSTVSTP 317
 QY 289 DEKNADVEYLNLANHDKVFPVANN---LDGSPANIFEGGEATSTTGKLVGIIKQGDYKVE 345
 Db 318 D--GNKVVVFSEKSKDAVTVVINGKEFTATP-----BENTVTLTADVASVKNGE-AFN 370
 QY 346 VQVTKRGGLTVGNT-----GIITVK-NLDTPASAIKNVPALDADNNGVNYGSKLSGKD 399
 Db 371 VIVTGAKDL-VGNTWEMVEGKATYKVEKDVTAPEVKDI--KVXELVDGVATLEVTSEE- 426
 QY 400 FALNSQNLVVGKASLNLKLVATIAGEDKVPDPGSI--KSNHGIIISVNNYITABAAGE 458
 Db 427 --LSAOGKVVVKKG-----DKVVEGATIALDQDPTTKALVTVPGALTDKETA-- 470
 QY 459 ATLTIK-VG--DVTKDVKFKVTTDSRKLVSVKANPDKLQVQVONKTLPTVPTVTTDQGDPF 515
 Db 471 ANLKVEFVGKDAANNVGNKVTKEVKTKDVVA-PNLVKVVDENKAATP-TFD----- 522
 QY 516 GANTAAIKVLEPKTGVAEGGLDVTTDSGIGTKTIGVTGNDVGG--TWHFQ-NGN-- 570
 Db 523 -----REV-----TAQEGKRLVINLTSKDVTKEVAVASVEDNKKAIITLTFQEGNYK 570
 QY 571 -GATLGSILVNVTEGN--VAF-KNPELVSKVGQVQSPDTKLD--LNVSTTVYVQLSKYT 624
 Db 571 VAATKG--FVKDTAGNESAAFTKEVVEK-----KEEGKDEVAPEKAKVERVADSKT 622
 QY 625 SRVVSDDPENLEGYEVES-----KNLAVADAKIVGNKVVTGTPKGVDIHL----- 671
 Db 623 KFTVTFDEKVGKGQGADSNVNNYTLAGAKLPEGLTIVVNDGKSVTIELPETTFEKS 682
 QY 672 -----TKNGATAGKATVEIVQETIAIKSVNFKPQVOTENFVKKINIGTVLELEK 720
 Db 683 ETVKFTVANVANKGVKGG--TTNLLVNVVDTKAPPEKSAKITKVDAKEITL-TFSEAVN 739
 QY 721 SNLDDIVKGINLTKTQHKVRVVKSGAQGLYLDRNGDAVFNA---GDVKL---GDVTVS 775
 Db 740 VDATDFVLDLNGVLTATK--ADETAEASK-----DVVLKVTPADVNLATGTATVK 789
 QY 776 -----QTSDSALPNFKADLYDTILT 794
 Db 790 AXEVEGKVVLNTADTSINKNKLVAFKPV 818

RESULT 2

Q9ZESS
 ID Q9ZESS PRELIMINARY; PRT: 816 AA.
 AC Q9ZESS;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE CTC protein.
 GN CTC
 OS *Bacillus thuringiensis*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 ON NCBI_TaxID=1428;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CTC;
 RA Sun M., Yu Z.;
 RT "S-layer protein gene of *Bacillus thuringiensis* CTC."
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ012290; CAA09981.1;

DR InterPro; IPR003343; Big_2.
 DR InterPro; IPR009964; Invasin_intimin.
 DR Pfam; PF02368; Big_2; 1.
 DR Pfam; PF00395; SLH; 3.
 DR SMART; SM00635; BID_2; 1.
 KW S-layer.

SQ SEQUENCE 816 AA; 87293 MW; 1002B92905802151 CRC64;

Query Match 17.4%; Score 730.5; DB 2; Length 816;

Best Local Similarity 30.6%; Pred. No. 6.4e-22;
 Matches 269; Conservative 119; Mismatches 321; Indels 169; Gaps 36;

QY 1 AGKSPDPVAGHMAEGSINYLVKGAITGKPDGTGYPTESIDRASAIVFTKIINLPVDE 60
 Db 30 AGKSPDPVADHWGIDSYLVKGAITGKPDGTGYPTESIDRASAIVFTKIINLPVDE 89
 QY 61 NAQSFQDAKNNINGSKYIAAEVAGVVGKGGKGFPEGKIDRASFASMLVSYNLDKV 120
 Db 90 DAKSFADSQQWTPPIAAVEKAGVTKGTG-NGFENGKIDRVSWASLLVEAYKLDTKV 148
 QY 121 NGELVTTFTEDLLD-HWGEKANILNLGIVSG-TGGKWEFNKSVSPAAQAQIALTDK 178
 Db 149 NGTPTATFKDLTLNMGKCANILVELGISVGTADKWEPKTVTKAQAQFLAKTDKQF 208
 QY 179 G-----KKDNACAYVT---DVKVSEPTKLTGTGLDKLSADVTTL-----EGDKAVA--TE 225
 Db 209 GTEVAKVESAKAVTTQKVEVFSK-----AVEKLTEDVKLANKANNDKVLVDKVK 259
 QY 226 ASTDGTSAVTLGGKVPANCDLTVKVK--NQSFVTFVVEVKKLAVEKLTFFDDRAGQAI 283
 Db 260 LSEDKSATVELYSNLAAKQTYTVDNVKGVEVTVGSLAEKTIEMADQTVVADEP-TAL 318
 QY 284 AFKLNDEKNADVEYLNLANHDKVFPVANNLDGSPANI-PEGGEATSTTGKLVGIIKQGDY 342
 Db 319 KITVDENGTEV-----SPAGIEFTVPAEKNKAGEITLAKGTS 359
 QY 343 KRVQVTKRGGLTVSNGTIITVKNLDTPASAIKN-VVFALDAD-----NDGVV--N 390
 Db 360 TTVKAVTKDQKGVKVAESKEVKSAGETAVASISNWTVAEAKADFTSKDFKQNDKVVEGN 419
 QY 391 YGSKLSGKDPALNSQNLAVGEEKASLNLKLVATIAGEDKVPDPGSIKSNHGIIISV--- 446
 Db 420 VSVQVELKDQNNVNNVVKAEYESLVEVAVV--DKAT--GKVTVLGAKAPKVTVKD 474
 QY 447 -----VNNYITABA-AGEATLTIKVGVDVTKVVKFVTTDSRKLVSVKANPDKLQVQNK 499
 Db 475 SKGKELVSKTVEIEAFAQAQKAMKEIKLEKTNVALSTKDVTDKF-----VKA----- 519
 QY 500 TLPVTFVTQYGDPPFGANTAAIKVLPKTVGVVAGGLDVVTTDSGSICTKTIIGVTGNDV 559
 Db 520 --PV-----LDQYGEFAAPVEV--KVLDKDGKELKEQKLVAKYENKEL---VLNAHQGEA 568
 QY 560 GEGTVHFQNGNGATLGLSYLVNVTGNAVAFKNFELVSKVGQYQSPD--TKLDLNVSTTVE 617
 Db 569 GKTYVELTAKSGKK-----EVSKLALAKAPGVFSKFDVR---GLE 607
 QY 618 YOLSKYTSDRVSDPENLEGYEVESKNLAVADAKIVGNKVVTGK-----TPGKVDIH 670
 Db 608 NELDKYVTEENKKNEMVSVLPDANGLVLRKEAATLKVTTTDDKGVVDAISQVAVN 667
 QY 671 -----LTKNGATAGKA-TVEIVQETIAIKSVNFKPQVOTENFVKKINIGTVLEKSNL 723
 Db 668 DAAGTITVGNKAGETKYKTVVADGKLITTHSFKVVVDTPAPAKK-----LAVDETSTSL 722
 QY 724 DDIVKGINLTKTQHKVRVVKSGAQGLYLDRNGDAVFNA---DVKL-----GDVTVS 775
 Db 723 NEVAQGSSEL-----KTALLN-----ILSDGVPTATAGATVDVKFVSADTNVISE 768
 QY 776 QTSDSALPNFKADLYDTILTITKYTDKTLVFKVLKQKDV 813
 Db 769 ETAKFGTKGSTSIFVKELTVK---KGEQTKQVELDKPV 803


```

QY 121 NGELVTFEDLLDHWGEEKANILINLGISVGTGKWEPPNKSVSRAEAQFIALTD----- 175
DB 143 IGEPTQPKDLEPHWGKQANILVALEISKGTGNGWNPGEVTVTRAEAAQFIAMADQNKTS 202
QY 176 --KKYKKDQAAY-----VTDVKVSEBPTKLTLAGTGLDKLSADDVILEGDKAVAIE 225
DB 203 TSKMYMNRNIVITHPSLSSGIDVQ-HKQWTEVT-----E 239
QY 226 ASTDGTSAVVT-LGSKVAPNKLTVKVK--NQSFVTKFVYEVVKLAVEKL-TFDD----- 276
DB 240 QRADGWLKIVTSGKEWTP--LTKETIINEEFT--YETASHSSKVLGTVNAQTVTV 293
QY 277 -DRACQATAFKLNDEKGNADVEYLNANHDVKFVANNLDGSPANIPFEGEATSTTGKLAIV 335
DB 294 MEESGSIRIRVAGAGFQWQDNQNLNFKQE-----NLEGGK-ALIIDPGHGGIDSGNGVY 347
QY 336 GIKQDYKVEQV-----TKRGGLTVSNTGIIIVK-NLDTPASAIKNVVFALDADNDGV 388
DB 348 YEKESETVLDVSLRLKIFEQKAPTVMFTRTDNRPGVNSTDSLKRVKVEFAQEHNGDIF 407
QY 389 VNYGSKLSGKDPALNSQNL-----VVGKASLNKLVIATIGEDKVVDPGSIKSSN 440
DB 408 VSIHANGSAEKNQGQETETLYQSAKVTNPNHVEDSKLLA-----QKIQRLLVAALGTGD 462
QY 441 HGISVWNNYITAEAGATITIKVGDVTQDKVFKVITDTSRKLVSVA 488
DB 463 RG-VKHQDLVYTRNTWPAVLTELAFVNDKSDADKIATPKORQAAEA 509

RESULT 5
Q81S57 PRELIMINARY; PRT; 414 AA.
AC Q81S57;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylmuramoyl-L-alanine amidase, family 3.
GN BAI817.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzaple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Knouri H.M., Radune D.,
RA Benton J.L., Mahancud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86 (2003).
DR EMBL; AF017029; AAP25724.1; -.
DR TIGR; BAI817; -.
DR GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002508; Amidase_3.
DR Pfam; PF01520; Amidase_3; 1.
DR SMART; SM00395; SLH; 3.
DR SMART; SM00646; Ami_3; 1.
DR Complete proteome.
SQ SEQUENCE 414 AA; 45351 MW; 40319D11B618F18F CRC64;

Query Match 13.2%; Score 556.5; DB 16; Length 414;
Best Local Similarity 39.2%; Pred. No. 3.1e-15;
Matches 145; Conservative 54; Mismatches 120; Indels 51; Gaps 14;

```

```

QY 3 KSPDPVPAGHWAEGSINYLVKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA 62
DB 27 KTFPDVQVTHWAEDSINYLAEGKAVTGNKGFEGKEITRAEATMMAKILNLPIDKNA 86
QY 63 QPSFDKAKNISSKVIAAVEKAGVVGKDGKFNPEGKI DRASFASMLVSAFNKDKVNG 122
DB 87 KFSYADSQKHATPIIAAVEKAGVVGKDG-NGFPDGI DRVSMASLLVEAVKLESKVG 145
QY 123 ELVTFEDLLD-HWGEKANILINLGISVGTGKWEPPNKSVSRAEAQFIALTDK-KYK 180
DB 146 TPATKFKOLETLNWKGEKANILVELGISVGTGDKWEPPKILTKAEAAQFIKADSLKVG 205
QY 181 KNAQAYVTDV-----KVSEPTKLTJ-TGTGLDKLSADDVILEGDKAVAI--EA 226
DB 206 PLVEKVIIDPGHGFDPNGFQGVSEKIVFDTSRLQKLEKNTPL---KALLTREEN 262
QY 227 STDGTSVAVTLGGKV-----APNKDLTVKV-----KQSFVTKFVYEVVKLAVEKLTFDD 276
DB 263 GNPGSNKNSLANRVKFGQENNADI FVSIHANSSEKHGHTGTETYYVKKSGEETQIEK 322
QY 277 DRAGQAIKLNDEKGNADVEYLNANHDVK-----FVANNLDGSPA-----NIFE 322
DB 323 D--SEVLAKKIQ-----NRVVEALHTRDKIKDDHSLYVNNN-NTVPAVLTELAFIDNDID 375
QY 323 GGEATSTTCK 332
DB 376 NGKLATESGR 385

RESULT 6
Q813K8 PRELIMINARY; PRT; 483 AA.
AC Q813K8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE S-layer protein / peptidoglycan endo-beta-N-acetylglucosaminidase
DE (EC 3.2.1.1.-).
GN BC2464.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapratral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Busch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91 (2003).
DR EMBL; AF017006; AAP09426.1; -.
DR GO; GO:0004040; F:amidase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002901; Amidase_4.
DR InterPro; IPR001119; SLH.
DR Pfam; PF01832; Amidase_4; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00047; LY22; 1.
DR Glycosidase; Hydrolase; Complete proteome.
SQ SEQUENCE 483 AA; 53801 MW; 13DOCCALF2911A6D CRC64;

Query Match 12.6%; Score 528.5; DB 16; Length 483;
Best Local Similarity 31.3%; Pred. No. 5.1e-14;
Matches 147; Conservative 70; Mismatches 145; Indels 107; Gaps 14;

QY 5 FPDVPAGHWAEGSINYLVKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAOP 64

```

```

Db 40 FPDVPT--WADKSNVYLQVNLNGVDPDGTGFSNDSLDASATKIMTKVLGIKIDNAPK 97
Qy 65 SFXDARNWSSKYIAAVEKAGVVGKDGKENVPECKIDRASFAFMLVSAYNLKDKNVNG 124
Db 98 SFTDSQNHATPYIAAEKAGIVKGEQNGIFNPSGKVTAAATMLVNAVYKLQSTANHNE 157
Qy 125 VTFPEDLDDHGEKANIILNIGISVGTGKWPENKSVSRAEAAQFIALTDDKYGKQNA 184
Db 158 QVRFEDLKGHWGKGANILIDLKISNGTENGWPNRFRITRAEAAQTAKTMDLQWLND 217
Qy 185 QAYVT-----DVKVSPTKLT-----LGTGGLDKLSADD-----VTLGDK 220
Db 218 KEVITATSYEDNLTVASKITAQEIIDSFIAXHSDSPLVGHGQDFINAQNGVSAHYLA 277
Qy 221 AVAIEASTDGTSAVV-----TLGCKVAPNKD----- 246
Db 278 AHALLGSGYKSEIAYOKHNLFLGRAYDGPFFKAKYLPYSGDSIAYNANYRERYLEES 337
Qy 247 -----LTVKKNQSFVTK-----FVYEVKLAVEKLTPDDDRAGQAIAFKLNDEK 291
Db 338 GMYNGSTLTGMNVKYASDKGWAKKIAGIMERIKPRVEDYTY-----AKLPKNP 388
Qy 292 GNADVEYLNLNHDVKFVANNLDGSPANIPEGGEATSTTCKLAVGKDGKYKVEQVTKR 351
Db 389 ETLDDVDAL---SNBIPY-KMYADGSSNV-----VSSAT-----YYQVPYPPNLIKSR 433
Qy 352 GGLFTVSTGTITVKNLDTASAIVNVFALDADNDGVNVYSGKSLGKDF 400
Db 434 PDVAENKVGTV---TPGTTI-----FIYREDPNGWVEFSFEANGKEY 474

RESULT 7
Q8KYE8 PRELIMINARY; PRT; 531 AA.
AC Q8KYE8
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Surface-layer N-acetylmuramoyl-L-alanine amidase, (pXO2-42).
GN BXE0045.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=A2012;
RX MEDLINE=22061436; PubMed=12004073;
RA Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L.,
RA Holtzapfel E., Busch J.D., Smith K.L., Schupp J.M., Solomon D.,
RA Keim F., Fraser C.M.;
RT "Comparative Genome Sequencing for Discovery of Novel Polymorphisms in
RT Bacillus anthracis."
RL Science 296:2028-2033(2002).
DR EMBL; AE011191; AAM26204.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002508; Amidase_3.
DR InterPro; IPR001119; SLH.
DR Pfam; PF01520; Amidase_3; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00646; Ami_3; 1.
DR Plasmid.
KW SEQUENCE 531 AA; 58860 MW; 45CBFFDD398DD7AC CRC64;
Query Match 12.6%; Score 528; DB 2; Length 531;
Best Local Similarity 33.2%; Pred. No. 6.1e-14;
Matches 158; Conservative 68; Mismatches 194; Indels 66; Gaps 17;
Qy 3 KSFDPVAGHABGSINVLVNGKAITGKPDGTYGTFESIDRASAAVFTKILNLPVDENA 62
Db 31 KTFTDVP--NNAQGSVNTLMKK-ALDGPDPGTFSFKIDRGSAAKLMNVGLQINKQA 87

```

```

Qy 63 OPSPKDANTWSSKYIAAVEKAGVVGKDGKENVPECKIDRASFAFMLVSAYNLKDKNVNG 122
Db 88 KPSFDQAKNHWASPYIAAVEKAGVVGKDGKENVPECKIDRASFAFMLVSAYNLKDKNVNG 147
Qy 123 ELVTFPEDLDDHGEKANIILNIGISVGTGKWPENKSVSRAEAAQFIALTDDKYGKQNA 182
Db 148 DLPTQFEDLKGHWGKGANILNIGISVGTGKWPENKSVSRAEAAQFIALTDDKYGKQNA 205
Qy 183 NAQAYVTD--VKVSEPTKLTLTGTGLDKLSADDVTLGDKKVAIEASTDGTSAVV--LG 238
Db 206 TSKMYMNRHFTTHQP-----SLSGVTSNQHAPQI-----IVKEQBADGWIKIVNIG 256
Qy 239 GKVAP--NKDLTVKKNQSFVT--KFVYEVKLAVEKLTPDDDRAGQAIAFKLNDEK-- 292
Db 257 DKWPLVKEKRETI--HSTFTTYPEASHSSKVLGT-----HSPQTVT--VIEEKGSW 303
Qy 293 -----NADVEYLNLNHDVKFVANNLDGSPANIPEGGEATSTTCKLAVGKDG--YKVE 345
Db 304 IRIRTNAGFQWLDKNQLTLPKQNNFLEGKTIIDPHGGIDGGHKGKGYMNEPVPVYDTA 363
Qy 346 QVTKRGGLTVSNTGIIITVKNLDTPA-----SAIKNVVVFALDADNDGVNV----- 390
Db 364 VRVQKLFAKTPFTTALLTRDAYSRGKNATDSLGRVEFAKKNKGDIFVSTHANGFNGA 423
Qy 391 -----YGSKLSGKDFALNSONLVVGEKASLNKLVAITAGEDKVDVDPGSIKSN 440
Db 424 HGTEFTFYKAPTQKSNPVNDSRIAEKIQ-KRLITALTQTRDRGVKIGNLYLVLEN 478

RESULT 8
Q8IU01 PRELIMINARY; PRT; 344 AA.
AC Q8IU01
DT 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE S-layer protein, putative.
GN BAI093.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapfel E.K., Okstad O.A., Helgason E., Rikstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
DR EMBL; AE017027; AAP25072.1; -.
DR TIGR; BAI093; -.
DR InterPro; IPR006635; NEA transcript.
DR InterPro; IPR001119; SLH.
DR Pfam; PF05031; NEAT; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00725; NEAT; 1.
DR Complete proteome.
KW SEQUENCE 344 AA; 38277 MW; 48764AC58FDD72241 CRC64;
Query Match 12.1%; Score 509; DB 16; Length 344;
Best Local Similarity 59.9%; Pred. No. 2e-13;
Matches 103; Conservative 22; Mismatches 45; Indels 2; Gaps 1;

```



```
QY 5 FPDVPAGHWAEGSINYLVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDENAQP 64
DB 169 FSDVPT--WAQESQVQLVDKEAIVGKPDGTGPTAPAESIDRSSAKILATVLRLEIKDKAKP 226
QY 65 SPKDAKNWSSKYIAAVEKAGVVGKDGKGFNFPYEGKIDRASASMLVSNYLNKDKVNGEL 124
DB 227 SFPDAQNHWPATPIAAVEKAGIVKDGKGFNFPYEGKIDRASASMLVSNYLNKDKVNGEL 124
QY 125 VTFPFDLLDHWGEEKANILINLIGISVGTGKWEKPNKSVSRABAAQFIATLTDK 176
DB 287 PKFADLNHNGAKYANILIQEIKISIGTDNGWAPNKAVSRABAAQFIACADK 338

RESULT 9
Q81QA4 PRELIMINARY; PRT; 410 AA.
ID Q81QA4
AC Q81QA4
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE N-acetylmuramoyl-L-alanine amidase, family 3.
CN BA2528.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22609414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Ostad O.A., Helgason E., Rillstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.B., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaitte J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RA "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86 (2003).
DR EMBL; AB017032; AAP26383.1; -.
DR TIGR; BA2528; -.
DR GO; GO:0008745; P:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002506; Amidase_3.
DR InterPro; IPR001119; SLH.
DR Pfam; PF01520; Amidase_3; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00646; Ami_3; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 410 AA; 45230 MW; 28F7F51302F92428 CRC64;

Query Match 12.0%; Score 506; DB 16; Length 410;
Best Local Similarity 34.1%; Pred. No. 3.4e-13;
Matches 141; Conservative 62; Mismatches 139; Indels 72; Gaps 15;

QY 5 FPDVPAGHWAEGSINYLVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDENAQP 64
DB 29 FPDVPA--WADKSVTYLVDKQVLSGYPDGTGFGSDTLDRAAATMTKALGIHIDLNAPK 86
QY 65 SPKDAKNWSSKYIAAVEKAGVVGKDGKGFNFPYEGKIDRASASMLVSNYLNKDKVNGEL 124
DB 87 SFTDSQNHWPATPIAAVEKAGIVKDGKGFNFPYEGKIDRASASMLVSNYLNKDKVNGEL 146
QY 125 VTFPFDLLDHWGEEKANILINLIGISVGTGKWEKPNKSVSRABAAQFIATLTDK 184
DB 147 QSKFEDLKGHWGEEKANILIDLKISVGTGNGWQPNKFTIRAEAAQLTAKTDMLOYSHNP 206
QY 185 QAYVT-----DVKVSPTKLT--TGGLDKLSDVDTLEGDKAVAEASTD 229
DB 207 LENKTIIDPGHGSDPGKDTKGLPESKIVLDTLSRLQKL-----LEKHTP 252
```

```
QY 230 GTSAVTLGSKVAPNKD-----LVTKV-----NOSFTVKFYEVKKLAVEKLTEDDDRAQ 281
DB 253 FT-VLLTRKSDTRPGHDQKSSLOERVKFAKQNGDIF-----ISVHANAFNGNAKGT 303
QY 282 AIAF--KLNDKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLVGIIKQ 339
DB 304 ETYIYKSSKEKINPVE-----ESRVLAEKIQIRLV-----DALQTRDR---GVKH 347
QY 340 GDYKVEVQVTKRGGLTVSNTGITVKN--LDTPSAIAKN--VFPALDADNDGVVNY 391
DB 348 GD-----LHVIRENDMPAVLTAFIDNGIDYSKLSTENGROIAAEATYEGILDY 397

RESULT 10
Q81AN7 PRELIMINARY; PRT; 410 AA.
ID Q81AN7
AC Q81AN7
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE S-layer protein / N-acetylmuramoyl-L-alanine amidase (BC
DE 3.5.1.28).
GN BC3524.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fongstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RA "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91 (2003).
DR EMBL; AB017009; AAP10458.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002508; Amidase_3.
DR InterPro; IPR001119; SLH.
DR Pfam; PF01520; Amidase_3; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00646; Ami_3; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 410 AA; 45571 MW; 318702F62DDE6341 CRC64;

Query Match 12.0%; Score 504; DB 16; Length 410;
Best Local Similarity 54.7%; Pred. No. 4.1e-13;
Matches 98; Conservative 27; Mismatches 52; Indels 2; Gaps 1;

QY 5 FPDVPAGHWAEGSINYLVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDENAQP 64
DB 29 FPDVPA--WADKSVTYLVDKQVLSGYPDGTGFGSDTLDRAAATMTKALGIHIDLNAPK 86
QY 65 SPKDAKNWSSKYIAAVEKAGVVGKDGKGFNFPYEGKIDRASASMLVSNYLNKDKVNGEL 124
DB 87 SFTDSQNHWPATPIAAVEKAGIVKDGKGFNFPYEGKIDRASASMLVSNYLNKDKVNGEL 146
QY 125 VTFPFDLLDHWGEEKANILINLIGISVGTGKWEKPNKSVSRABAAQFIATLTDK 183
DB 147 QVKFEDLKGHWGEEKANILIGLISVGTGNGWQPNKFTIRAEAAQLTAKTDMQHRQKN 205

RESULT 11
Q81HB4 PRELIMINARY; PRT; 530 AA.
ID Q81HB4
AC Q81HB4
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
```

```
01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE S-layer protein / N-acetylmuramoyl-L-alanine amidase (EC
DE 3.5.1.28).
GN BC0902.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatral V., Bhattacharya A., Reznik G., Mikhailova N., Iapichus A.,
RA Chu L., Mazur M., Goltseman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Kuych G., Hasekorn R., Fongstein M., Ehrlich S.D.,
RA Overbeek R., Pychides N.,
RA "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis";
RL Nature 423:87-91(2003).
DR EMBL; AE017000; AAP07889.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008745; P:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro; IPR002508; Anidase_3.
DR InterPro; IPR001119; SLH.
DR Pfam; PF01520; Anidase_3; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SMO0646; Ami_3; 1.
DR PROSITE; PS01072; SLH DOMAIN; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 530 AA; 58834 MW; 37E448FB109D2598 CRC64;

Query Match 11.3%; Score 475; DB 16; Length 530;
Best Local Similarity 30.3%; Pred. No. 6.7e-12;
Matches 150; Conservative 79; Mismatches 210; Indels 56; Gaps 19;

QY 3 KSFPDVPAGHWAEGSINYLVDKGAITGPKDGTGPTESIDRASAIVFTKILNLPVDNA 62
DB 27 KKFPDVSANSKDAIYLVVERNVINGPDPNGFPGYGLNTRAQAKIATAIGAKVDNA 86
QY 63 QPSFKDANWSSKYIAAVEKAGVVGKGNFPEPKIDRASPASMLVSNLKDQVNG 122
DB 87 KPSYDANKNSWAASFAIAAMEKENIKGREPGVPFPEGKVTAEAAWMLVRAYNLKSQV 146
QY 123 ELVATTFEDLLHGWGEKANILNLGIVSGTGKWPENKSVSRABAAQFIATDKKYGKD 182
DB 147 PVPYKFALENHGWKEEVNVLKSLGTENGKPNDSITREQAQLTAQTD-KFSK-- 203
QY 183 NQAAYVDVKSSEPTK-LTLGTGLDK-LSADDVTLEGDKAVAIEASTDGTSAVTLGG- 239
DB 204 NSDRPVETKMYIDRKFTYHAPSLSGSIANSQ---HNPQVTEIKERDQGWIKIATSGD 260
QY 240 KVAENKDLTVKKNQSFVTKFYVEKVLAVEKLTFTDDDRAGQALAFKLND----KGNAD 295
DB 261 KWTPLVEKT-EVINEGFTT-----YAEASSSKWNGTHNAQVTVRENGSWIRWGAG 314
QY 296 VEYLLANLHNDVKNVANNLDGSPANIFEGGEATSTTGKLVAGIKQG-----DYKVEVQ--VT 349
DB 315 FQWKNKNQNPVKQGNFLEKQ-AIIIDPHGGVDPHSGVKNDESAILVLTSLRVQKLFE 373
QY 350 KKGGLTVSNIGTIIVKNLDTPASAT-KNVVFDALDADNGVN-----YGSKLSGKD---- 399
DB 374 QKTPTVLLTRNDTRPNTGPKESLKKRVEFAQENKGDIFVSIHANGNEQVGETTFY 433
QY 400 -FALN---SQNLVGEKASLNKLVIATIAGEDKVDPGSISIKSSNHGIIISVNV----- 448
DB 434 RSATPNSEESVLAEKVQ-KELVQALQSNDR-----GVKTENFYVVKYNTWPAILAE 485
QY 449 -NYITAEAAAGEATLT 462
DB 486 LGFIDAKGEGEKIAT 500

RESULT 12
```

```
Q9L655
ID Q9L655; PRELIMINARY; PRT; 931 AA.
AC Q9L655;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Surface layer protein.
GN SLPA.
OS Bacillus pseudofirmus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=79885;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OF4; PubMed=11029415;
RA Gilmour R., Messner P., Guffanti A.A., Kent R., Scheberl A.,
RA Kendrick N., Krulwich T.A.,
RA "Two-Dimensional Gel Electrophoresis Analyses of pH-Dependent Protein
RT Expression in Facultatively Alkaliphilic Bacillus pseudofirmus OF4
RT Lead to Characterization of an S-Layer Protein with a Role in
RT Alkaliphily.";
RL J. Bacteriol. 182:5969-5981(2000).
DR EMBL; AF242295; AAF68436.1; -.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
SQ SEQUENCE 931 AA; 96855 MW; 6A9727171COA78D0 CRC64;

Query Match 9.4%; Score 395; DB 2; Length 931;
Best Local Similarity 24.2%; Pred. No. 3.2e-08;
Matches 233; Conservative 106; Mismatches 336; Indels 288; Gaps 43;

QY 1 AGKSPDVPAGHWAEGSINYLVDKGAITGPKDGTGPTESIDRASAIVFTKILNLPVD- 59
DB 34 ADAKFSQVSSSHWADDINYLVEKAIQGYDGTGPKPNSITRAIIAVLANTLGLDVS 93
QY 60 ENAQSFQDA-KNIWSSKVIIV--EKAGVVGKDGKEN--FYPEGKIDRASPASMLVSA 114
DB 94 EVTTDKFSDVPATHWANFYIAIVDQTEGVI--DGYNGTFPSPNTITRQEVAKWVVEAY 151
QY 115 NLKDKVNGSLV---TTFEDLLHGWGEKANILNLGIVSG--TGGKWPENKSVSRABAA 169
DB 152 DL-----ELVEGKDLPTFDSGLASTDYINLASNGVAGWATTTAFPRGEVLRAQTA 205
QY 170 FI--ALTDKXKKNQNAQAYVTDVVKVSEPTKLTLT-----GTGLDKLSADDVTL 216
DB 206 FIIHRAVEERIEVPGLVPTVSSVAVNATLTTLVVSDDTTHVTLDTALVANEATEVTF 265
QY 217 E-----GDKAVATEASTDGTSAVVTLGGKVPAPNKDLTVKKNQSFVTKFY 262
DB 266 EINDVEYTVTWVVEELAVASVNAVNGQLVNFPTQPDASSVLTSPTASTVTKTGV 325
QY 263 EVKCLAVEKLTFTDDDRAGQALAFKLNDKGNADVEYLLANLHNDVKNVANNLDGSPANI 322
DB 326 SLNR-----TSQDTHSNDW-----LVGAERAVL-----SSDGKTLTV-- 358
QY 323 GGEATSTTGKLVAGIKQGYKVEVQVTKEGGLTVSNIGTIIVKNLDTPASAIKNVVVFD 382
DB 359 ----TAPTQGF-----KGNVDVTVDKAGSQNTIPSV-IKTSVDDTTAPIVNSVTVFA 409
QY 383 ADNDGVV--NYGSKLGGKDFALNSQNLVVGEGA---SLNKL-----VAT-----IA 423
DB 410 TDKFEVTLSEPIDSLGVEVLRLNGQPVASGFDALTGPTNKLTPARPSSVATGNTATYIA 469
QY 424 G-----EDKVDVDPGSIKSSNHG 442
DB 470 GFSDAAGNFVTPSTTTVPVTDTTALAVASLQVNTQKVRLLTFNKLNSASKTALSQSG 529
QY 443 II-----SVNNY-----ITAEAAAGEATLTIKVGDVT-----KDVFKVT 477
DB 530 VVTRNGSTTSNFTVAQNLTVDNTCK-TYDITLSDATVNGNSEVFGITLIKDAFTDVT 588
QY 478 TDSRKLVS-----VKANPKLQVQNKLTLPVTFVTTDQYDGPFGANT 519
```

Db 589 NNKNDLYSKSITLTKDVTAPVTTGAALASNRQAEVLTVSEGVTTIT-----NP 635
 Qy 520 AAIEVLPLKTC-----VVAEGGLDVVTTDSGSGIKTKTIGVTGN-DVGGTGVHFQNG-- 569
 Db 536 AQVK--LRKGAQGTGLSVALKGGTDNLVIVSYSGAELAGGSQVRLAAGAVTDLNGNA 693
 Qy 570 -NGATLSGLSYNNVTEG--NVAFN-----FELVSKVQYQSGSPDTKLDLNVSTTVVEYQL 620
 Db 694 NNAVNPASVSVSATPAAPLNAVAVANGTNTVFTVAPTQ----- 732
 Qy 621 SKYTSDRVYSPENLEGVESKNLAVADAKIVGNKVVTGKPGKVDIHL-----TKN 674
 Db 733 -TFTFASLNHNFKIDGQAVSN-----SDITLNSTRDIIIVSLPSEDSVKISGNALFTTN 787
 Qy 675 G-----ATAGKATVEIVQETIAIKSVNFKVPQVQTFNFEKKINTGTVLELEKSNLD- 724
 Db 788 GLAESGRALATA-TATVTVTDNT-----APTLTG--AQIVSANVIKFTFDENLDA 835
 Qy 725 -----DIVKGINLTQKHQVRVVKSGAEGQKLYLDNRGDAVFNAGDVKLGDDTVTSQT 777
 Db 836 LTLTDAADLDDIQLSNGTVAYNGGFASGATSG-----GDTVSSVDGKSLVVTVSPN 888
 Qy 778 SDS 780
 Db 889 SDS 891
 RESULT 13
 Q45664 PRELIMINARY; PRT; 920 AA.
 AC Q45664;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE SBSB Gene (Sequence 5 from PATENT WO9908567 precursor).
 GN SBSB.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PV72;
 RX MEDLINE=971197531; PubMed=9045827;
 RA Kuen B., Koch A., Asenbauer E., Sara M., Lubitz W.;
 RT "Molecular characterization of the Bacillus stearothermophilus PV72 S-
 layer gene sbsB induced by oxidative stress.";
 RL J. Bacteriol. 179:1664-1670(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PV72;
 RA Lubitz W., Resch S.;
 RT "SECRETION OF CARRIER-BONDED PROTEINS INTO THE PERIPLASMA AND THE
 RT EXTRACELLULAR SPACE.";
 RL Patent number WO9906567-A/5, 11-FEB-1999.
 DR EMBL; X98035; CA66724.1;
 DR EMBL; AX000222; CAB77070.1;
 DR InterPro; IPR003343; Big_2;
 DR InterPro; IPR001119; SLH.
 DR Pfam; PF02368; Big_2; 2.
 DR Pfam; PF00395; SLH; 3.
 DR SMART; SM00635; BID_2; 2.
 KW S-layer; Signal.
 FT CHAIN 32 920
 SQ SEQUENCE 920 AA; 97916 MW; 1F3C20344B40F3A2 CRC64;

Query Match 8.7%; Score 367; DB 2; Length 920;
 Best Local Similarity 23.3%; Pred. No. 4.4e-07;
 Matches 227; Conservative 138; Mismatches 369; Indels 240; Gaps 48;
 Qy 4 SPFDVPAGHWAGSINYLYDKAITKPKDGTGYPTESIDRASAAVIFTKILNLPVDENAQ 63
 Db 33 SFTDVAPOY--KDAIDFLVSTGATKGTETKFGVDEITRLDAVILARVLKLDVDAK 90

Qy 64 PSFKDAKNWSKYIAAVEKAGVVKGCKENFYPEGKIDRASPSMLVSAINLKDKNCE 123
 Db 91 AGFTDVPKD--BAKYNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKJK--ADD 146
 Qy 124 LVTTTEEDLLDHWGEBKANILNLGIVG--TGCKPEPKNSVSRABAAQFIALTDRKKYKGD 182
 Db 147 VKLPFTDNDTWA--PYVKALYKYEVTKGKTPTSFGAYQNIITRGDFAQFVY-----RAV 198
 Qy 183 NQAQVTVQVSEPKLTLTGTLGDKLSADDVT--LEGDKAVAI EAST--DGTSAVVTLG 238
 Db 199 NINAVEIVEVTAVNSNTVKVTFNTQIADVDFTNFADNGLTFTKATLSRDKSEVVVN 258
 Qy 239 GKVAENKDLTVK--VKN-----QSFVTKFVYEVKLAVEKLPFDDDRAGQAIKFLND 289
 Db 259 KEPTNRQETITATGINKLKGETAKELTGKFWVSQDATTVALNNSSLKVGESGLTVKD 318
 Qy 290 EKNADVEYLNANHDVKFVANNDGSPANI--PEGGEATSTTGKLAGVIGQGYKVEVQ 347
 Db 319 QDQK-----DVVGAKVELTSSNTNIVVWSSGEVSSAAKV--TAVKPGTADVTAK 366
 Qy 348 VTKRGGLTVSNTGIIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNL 407
 Db 367 VTLPGVLTNTFKVTVTEFV--OVNQGFLL--VDN-----LSNA-----PQNT 408
 Qy 408 VVGEKASLNKLVATTIAGEDKVY-----DPGS-----ISIKSNHGIISVNN-- 449
 Db 409 VAFNKA--EKVTSMEAGETKTVMYDTKNGDPETKPFVDFKDATVRSNLPITATAINGSE 466
 Qy 450 -YITAEA--AGEATLTIKUGDVK-----DYK-----FKVTTDSRKLVSKANPDKQ 494
 Db 467 LLVTANAGOSKASPEVTFKONTKRTFTVDVKQDLPQDIKVDATSVKLSAEVGGGEVE 526
 Qy 495 VVQNKTLPTVFTVTTDQYDGFPGANTAAIKEVLPTGTVVAEGGLDVTTDSGIGTKTIGV 554
 Db 527 GYNQTIKVS AV--DQYG-----KEI--KFGT--KGKVTVTNTTEGLVIKNVN-- 568
 Qy 555 TGNVGEGETVHFQNGNGATGLSLYV-----NVTEGNVAFKNFELVS----- 595
 Db 569 -----SDNTIDFDGNSATDQFVVVATKDKIVNGKVEYKFKNAGSDTPTSTKTITVNVV 623
 Qy 596 -----KVGQYQSGPDTKLDLNVSTTVVEYQLSKTSDRVYSDPENLEGYEYES-----K 643
 Db 624 NVKADATPVGLDIVAP--SEIDVNPANTAS-----TADVDFINFESVEIYTLDSNGRLK 676
 Qy 644 NIAVADAKIVGNK--VVVTGK-----TPGKVDIHLTKNGT-----AGK 680
 Db 677 KVTPTATTLVGTNDYVEVNGVLOFKGNDLTLTSSSTVNVVDVADGTTKPIPVKINS 736
 Qy 681 ATVEIVQETIAIKSVNFKPVQTE--FVEKKINIGTVLELEKSNLDDIVKGINLTKETQ 737
 Db 737 ASVP--ASAIVATSPVTVKLNSSDNDLTFFELIFGVIDPTQLVKD--EDINEFIASKA 793
 Qy 738 H-----KVRVKSGBE-----QKLYLDNR-----GDAVFN-- 763
 Db 794 NDGYLYNKPLVTVKDAAGKVIPTGANVYGLNHDATNGNIWFDEEQAGLAKKTSDFHFDVD 853
 Qy 764 --AGDVKLGDVTSQTSLSALPNFRADLYDLTITKTKDG-----TLVFK--VLKDKDV 813
 Db 854 FSLANVVKTGSGTVSSS-----PSLSDAI--QLTNSGDAVSEFTLVKISYVKGADK 902
 Qy 814 ITSEIGSQAVHNVV 827
 Db 903 DDNNLLAAPVSVNV 916
 RESULT 14
 Q9RER7 PRELIMINARY; PRT; 1268 AA.
 AC Q9RER7;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Surface layer protein precursor.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:19:59 ; Search time 62 Seconds
(without alignments)
3796.162 Million cell updates/sec

Title: US-09-844-281-1
Perfect score: 4202
Sequence: 1 AGKSPDPVAGHWAEGSINY.....ITSEIGSQAVHVNLPML 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4202	100.0	833	5	AAM47757 Mature EA
2	719	17.1	785	4	AAB84651 Amino aci
3	564.5	13.4	529	6	ABU17672 Protein e
4	556.5	13.2	414	6	ABU17636 Protein e
5	553.5	13.2	525	6	ABU18750 Protein e
6	459	10.9	379	6	ABU17384 Protein e
7	369	8.8	408	6	ABU18670 Protein e
8	369	8.8	1252	2	AAR80530 B. sphaer
9	367.5	8.7	921	2	AAR22863 Bacillus
10	367.5	8.7	921	3	AAB10626 B. stearo
11	367	8.7	920	2	AAR93253 B. stearo
12	257.5	6.1	1338	2	AAR41731 High mole
13	257.5	6.1	1598	2	AAR30291 Non-typea
14	252.5	6.0	941	2	AAR77395 Full leng
15	251.5	6.0	941	2	AAR07478 Cellulase
16	250.5	6.0	1601	2	AAR30292 Non-typea
17	249.5	5.9	1529	2	AAR41732 High mole
18	249.5	5.9	2353	2	AAR9393 Haemophil
19	246	5.9	1180	3	AAB01845 Haemophil
20	246	5.9	1188	3	AAB01844 Haemophil
21	242	5.8	1221	3	AAB01825 Haemophil
22	242	5.8	1227	3	AAB01824 Haemophil
23	240.5	5.7	2411	3	AAB23860 Haemophil
24	235	5.6	1004	3	AAB01841 Haemophil
25	235	5.6	1010	3	AAB01840 Haemophil

26	235	5.6	2712	6	ABU39146 Protein e
27	234.5	5.6	2314	4	AAB69136 M. catarr
28	231	5.5	1268	7	ADC00996 Enteroha
29	230.5	5.5	2154	6	ABU38771 Protein e
30	230	5.5	1325	4	AAG98256 Escherich
31	229	5.4	969	3	AAB01827 Haemophil
32	229	5.4	975	3	AAB01826 Haemophil
33	229	5.4	1612	5	ABB47333 Listeria
34	227.5	5.4	1073	3	AAB01837 Haemophil
35	227.5	5.4	1079	3	AAB01836 Haemophil
36	227	5.4	2504	6	ADA34534 Acinetoba
37	224.5	5.3	1386	6	ABB82573 H. influe
38	224.5	5.3	1449	6	ABB82570 H. influe
39	223.5	5.3	852	6	ABU15220 Protein e
40	222.5	5.3	2053	4	AAB69135 M. catarr
41	222	5.3	1992	2	AAR04505 Moraxella
42	222	5.3	1992	4	AAB69137 M. catarr
43	222	5.3	1992	4	AAB69133 M. catarr
44	222	5.3	2047	4	AAB69134 M. catarr
45	222	5.3	2265	6	ABU17199 Protein e

ALIGNMENTS

RESULT 1
AAM47757
ID AAM47757 standard; protein; 833 AA.
XX
AC AAM47757;
XX
DT 25-FEB-2002 (first entry)
XX
DE Mature EA1 protein.
DE Anthrax; antibacterial; vaccine; EA1 antigen.
XX
OS Bacillus anthracis.
XX
PN WO200183561-A2.
XX
PD 08-NOV-2001.
XX
PF 30-APR-2001; 2001WO-US013648.
XX
PR 28-APR-2000; 2000US-020050SP.
XX
PA (TETR-) TETRACORE LLC.
XX
PI Mangold BL, Aldrich JL, O'Brien TW;
XX
DR WPI; 2002-055457/07.
XX
PT Novel monoclonal antibody, useful for detecting B.anthraxis, and for
treating B.anthraxis infection, is specifically reactive against Bacillus
anthracis and is non-reactive with B.thuringiensis and B.cereus.
XX
PS Claim 6; Fig 1; 27pp; English.
XX
CC The present invention relates to a monoclonal antibody which is
specifically reactive against Bacillus anthracis EA1 antigen. The present
sequence is the mature EA1 protein from Bacillus anthracis. The EA1
protein corresponds to the eag gene. The monoclonal antibody is highly
specific for Bacillus anthracis, and can distinguish Bacillus anthracis
from closely related non-pathogenic species. The present sequence is
useful for producing a vaccine against Bacillus anthracis and the
antibody is useful for treating, preventing or controlling Bacillus
anthracis infection (anthrax)
XX
SQ Sequence 833 AA;
Query Match 100.0%; Score 4202; DB 5; Length 833;
Best Local Similarity 100.0%; Pred. No. 2.2e-238;

Matches	833;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	AGKSPDPVAGHWAEGSINYLVDKGAITGKPDGTGPTESIDRASA	60						
Db	1	AGKSPDPVAGHWAEGSINYLVDKGAITGKPDGTGPTESIDRASA	60						
QY	61	NAQPSFKDAKNIWSKYIAAEKAGVVGKDGKGFYPEGKIDRAS	120						
Db	61	NAQPSFKDAKNIWSKYIAAEKAGVVGKDGKGFYPEGKIDRAS	120						
QY	121	NGELVTTTFEDLLDHWGEKANILNLGTSVGTGKWEPNKSVSRA	180						
Db	121	NGELVTTTFEDLLDHWGEKANILNLGTSVGTGKWEPNKSVSRA	180						
QY	181	KDQAQYVTDVKVSEPTKLTITGTGLDKLSADDVTLBGDKAVA	240						
Db	181	KDQAQYVTDVKVSEPTKLTITGTGLDKLSADDVTLBGDKAVA	240						
QY	241	VAPNKDLTVKVNQGSFVTKFYVEVKGLAVEKLTDFDDDRAGQAI	300						
Db	241	VAPNKDLTVKVNQGSFVTKFYVEVKGLAVEKLTDFDDDRAGQAI	300						
QY	301	LANHDVKFVANLDDGSPANIPEGGEATSTTGKLVAGIKQGDY	360						
Db	301	LANHDVKFVANLDDGSPANIPEGGEATSTTGKLVAGIKQGDY	360						
QY	361	IITVKNLTPASAIKNVVFALDADNDGVNYSKLSGKOPALNSQ	420						
Db	361	IITVKNLTPASAIKNVVFALDADNDGVNYSKLSGKOPALNSQ	420						
QY	421	TIAGEDKVVDPGSIKSNHGIISVNNYITAAAGEATLTIVKGD	480						
Db	421	TIAGEDKVVDPGSIKSNHGIISVNNYITAAAGEATLTIVKGD	480						
QY	481	RKLVSVKANPKLVQVQNTLTPVFTTDDQYDGPFGANTAAIK	540						
Db	481	RKLVSVKANPKLVQVQNTLTPVFTTDDQYDGPFGANTAAIK	540						
QY	541	TTDSSIGTKITGTVGNDVGGTGHFQNGNGATLGLSVNVN	600						
Db	541	TTDSSIGTKITGTVGNDVGGTGHFQNGNGATLGLSVNVN	600						
QY	601	GQSPDTKLDLNVSTTVYQLSKYTSRDVYSDPENLEGYEVES	660						
Db	601	GQSPDTKLDLNVSTTVYQLSKYTSRDVYSDPENLEGYEVES	660						
QY	661	GKTPGKVDIHLTKGATAGATVEIVQETIAIKSVNFKPVQTE	720						
Db	661	GKTPGKVDIHLTKGATAGATVEIVQETIAIKSVNFKPVQTE	720						
QY	721	SNLDDIVKGINLTKETQHKVRVYKSGAQGLYLDRNGDAVF	780						
Db	721	SNLDDIVKGINLTKETQHKVRVYKSGAQGLYLDRNGDAVF	780						
QY	781	ALPNFKADLYDTLTYYTKTDKGLTVKVLKDKDVTSEIGSQ	833						
Db	781	ALPNFKADLYDTLTYYTKTDKGLTVKVLKDKDVTSEIGSQ	833						
RESULT 2									
ID	AAB84651	standard; protein; 785 AA.							
XX	AC	AAB84651;							
XX	DT	05-SEP-2001 (first entry)							
XX	DE	Amino acid sequence of a bacterial surface array protein (SAP).							
XX	XX	Surface array protein; SAP; bacterial detection.							
XX	XX	Bacillus anthracis.							

FH	Key	Location/Qualifiers
FT	Misc-difference 1..2	/note= "the nucleotides encoding these residues are not given"
FT		
FT		
PN	WO200149823-A2.	
XX		
PD	12-JUL-2001.	
XX		
PF	04-JAN-2001; 2001WO-US000358.	
XX		
PR	06-JAN-2000; 2000US-0174901P.	
XX		
PA	(BIOS-) BIOSITE DIAGNOSTICS INC.	
XX		
PI	Lee BA, Flores BM, Valkirs GE;	
XX		
DR	WPI; 2001-418358/44.	
DR	N-PSDB; AAH28330.	
XX		
PT	Novel methods and kits for detecting the presence of Bacillus anthracis in a test sample.	
XX		
PS	Claim 2; Page 59-60; 62pp; English.	
XX		
CC	The present sequence represents a bacterial surface array protein (SAP). SAP is used in the method of the invention. The specification describes a method for detecting the presence of Bacillus anthracis in a test sample. The method comprises contacting the sample with a capture reagent and detecting whether the a SAP is bound to the capture reagent, which is indicative of the presence of B. anthracis in the sample. The method is useful for detecting the presence or absence of B. anthracis in a test sample.	
XX		
SQ	Sequence 785 AA;	
Query Match 17.1%; Score 719; DB 4; Length 785;		
Best Local Similarity 29.6%; Pred. No. 8e-34;		
Matches 233; Conservative 123; Mismatches 289; Indels 142; Gaps 29;		
QY	1	AGKSPDPVAGHWAEGSINYLVDGKAITGKPDGTGPTESIDRASA
Db	1	AGKTPDPVADHWGDSINYLVEKAVGNDKGMFPGKELTFAEAATMMAQILNLPDK
QY	61	NAQPSFKDAKNIWSKYIAAEKAGVVGKDGKGFYEGKIDRAS
Db	61	DAKPSFADSQGQWYTPFIAAEKAGVVGKDGKGFYEGKIDRAS
QY	121	NGELVTTTFEDLLD-HWGEKANILNLGTSVGTGKWBPNKSVSRA
Db	120	NGPPTATKFDLETLNWKGEKANILVELGSLVGTGQWEPKVTYKAEAAQFIATDKQFG
QY	180	----KKDNAQAVT---DYKVSEPTKLTITGTGLDKLSADDVTL---EGDKAVAEIA--
Db	180	TEAAKVESAKAVTTQKVEVKPSK-----AVEKLTKEIKVTNKANNKNDKVLKEVTL
QY	227	SDTGTSAVTLGGKVPAPNKDLTVK-----KNQSFYTKFYVEVKGLAVEKLT
Db	231	SEDKRSATVELYSNLAAKQTYTVDNVKGVKTEVAVGSLAEKTIEMADQTVVADEP---TA
QY	283	IAFKLNDEKGNADVEYINLANHDVRFVANNLDDGSPANIPEGGEATSTTGKLVAGIKQGDY
Db	288	LQFTVKDENGTEV-----SPGIEFV-----TPA-----AEKINAKGEITLAKGTS
QY	343	KVEVQTVKRGGLTVSNTGIIITVKNLTPASAIKNVVFALDADNDGVNYSKLSGKOPAL
Db	330	TTVKAIVYKDKGVKVAESKEVKVSAEGAAVASISNMTVAEQNKAD-----FTSKDFKQ
QY	403	NSQNLVVGEKASLNKLVAIAGDKVDPGSIKSNH--GIISVNNYITAEAAAGEAT
Db	382	NNK-VVEGDNAVYQ---VELKQDFNAVTTGKVEYESLNTVEAVVDKATKVTIVLSAGAP
QY	461	LTIKVGD-----VTKDVKFVTTDSRKLVSVKANPKLVQVQNTL

Db 438 VKYTVKSGKALVSHHTVEIEAFQAQKAMDKLEKTWA---LSTKDVTD-----L 485
QY 502 PVTFTTDDQYDPPGANTAAIKVLPKGTG-VVAEGGLDVTTDSGIGTKTIGTGNVGV 560
Db 486 KVKAPVLDDQYGEFTAPVTV--KVLDDKGLKEQKLEAKYVARELV-----LNAAGOEAG 539
QY 561 EGTVHF--QNGGATGLSLVYVTEGVAFKFNELVSKVQYQSGSPDTKLDLNVSTTVEY 618
Db 540 NYTVVLTAQSGEKEAKATLAL--KAPGAFSKEV-----RGLDTELD----- 581
QY 619 QLSKYTSRVSYPDENLEGVEVESKNLAVADAKIVGNKVVTGKTPQKVD----- 668
Db 582 ---KYVTEENQKAMTVSLPVDANGLVLKGAEEALKVTTNKKEGKENDATDAQVTVQN 638
QY 669 --IHLTKNGATPAGKA-TVEIVQETIAIKSVNFKRQVQENFVKKINIGTVELEKSNLDD 725
Db 639 NSVITVGGQKAGETKYKVVVLDGKLITHTSFKVVDVTAFAK-----GLAVETSTSLKE 693
QY 726 IVKGINL 732
Db 694 VAPNADL 700

RESULT 3
ABU17672
ID ABU17672 standard; protein; 529 AA.

AC ABU17672;
DT 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #3199.
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW Bacillus anthracis.
OS WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA21542.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 45596; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 529 AA;

Query Match 13.4%; Score 564.5; DB 6; Length 529;
Best Local Similarity 32.6%; Pred. No. 5.7e-25;
Matches 172; Conservative 81; Mismatches 192; Indels 83; Gaps 18;
QY 1 AGKSPDPVAGHWAEGSINYLVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDE 60
Db 25 AAKKFSVDPT--WAQGSVDYLVGKALDGPDTGPTSPSEAVDRGSAAKILAVVLGLPVDP 82
QY 61 NQAPFKDAKNTSSKYTAAYEAKGVKGDGKENVYBEGKIDRASFASMLVSAYNLKDVK 120
Db 83 KAKPSFKAQNHAAPIYAAVEKAGVINGDGTGKFNPSQINRAVSWLVAQYSLDKKI 142
QY 121 NGELVTFPEDLLDHWEKCANILINLIGISVGTGGKWEFNKSVSRAAAQFIALTD----- 175
Db 143 IGEFLTQFKDLFPHGKQANILVALEISKGTGNGWNPETVTRAEAAQFIAMADQNKTS 202
QY 176 --KRYGKKDAAQY-----VTDKVSSEPKLITLTGTLKLSADDVTLGSKRAVAIE 225
Db 203 TSKRMVMMRNVTYHQPSSLSGITDVQ--HKPQWVEVT-----E 239
QY 226 ASTDGTSAVVT--LGGKVAENKDLTVKVK--NOSFVTKFVYEVKLAVERL--TFDD----- 276
Db 240 QRADGWLKIVTSKGEKWT--LTKETINEEFTT---YETASHSSKVLGTYNQATVTV 293
QY 277 -DRAGQAIAPKLNDEKGNADVEYLNANHDVKFVANNLDGSPANIFEGEATSTTGKLA 335
Db 294 MEESGSWIRIRVAGGFQWVDKQNLNPFVKE-----NFLEK--AIIDPGHGIDSGNVGY 347
QY 336 GIKQGDYKVEVQV-----TRGGLTVSNTGIITVK--NLDTPASAIXNVVFPALDADNCGV 388
Db 348 YEKESETVLDSLRLLKIFEQAPPTVMTFRDTRPGVNSTDSLKKRVEFAQEHNGDIF 407
QY 389 VNYGSKLSGKDFALNSQNL-----VYGEKASLNKLVAIAGEDKVVDPGSIKSN 440
Db 408 VSHANGSAEKNGQGTETLYQSARAKVTNPHVDSKLLA-----QKIQDLVAALGTGD 462
QY 441 HGIISVNNYITAAAGENTLTKVGDVTKDKVKFVTTDSRLKLSVKA 488
Db 463 RG-VKHQDLYVTRENTMPAVLTAPVDNKSADKIATPKQQAAREA 509

RESULT 4
ABU17636
ID ABU17636 standard; protein; 414 AA.
XX
XX ABU17636;
XX
DT 19-JUN-2003 (first entry)

CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 525 AA;

Query Match 13.2%; Score 553.5; DB 6; Length 525;
 Best Local Similarity 31.8%; Pred. No. 2.5e-24;
 Matches 164; Conservative 71; Mismatches 172; Indels 109; Gaps 15;
 QY 1 AGKSPDPVAGHWAEGSINLVKGAITGKPDGTGPTSIDRASAAVFTKILNLPVDE 60
 Db 20 AAKKFSVPT--WAQSVLYVKGKALDGGKPDGTGPTSPSEAVDRGSAKTLAVGLPVPD 77
 QY 61 NAQSPFKDAKNVSSKIAAVERAGVYKGGKGFENYPEGKIDRASVSMVSNYNLKDVK 120
 Db 78 KAPSFDAQHNWAAPIAAVERAGVINGDTGKFNPSQINRASVSMVSNYNLKDVK 137
 QY 121 NGELVTTFDLDHWEKEKANILINIGISVGTGKWEKPNKSVRAAQAFTALTKYK 180
 Db 138 IGEPTQKDLPHWGGKQANILVALEISKGTGNGWNPETVTRAAQAQFIAMAERKIQ 197
 QY 181 KDAQAQ-----VTV-----KVSEPKLTLTGTDGLKSA----- 211
 Db 198 STKRMWNTNVTYHPSLSSGITDVHHPQNGKVTQ-----RADGWKMLTSREKW 252
 QY 212 DVTLEGDKAVAEASDTGTSVAVTLGGKVPAPKDLTVKYNQSFVTKFYEVKFLAVEK 271
 Db 253 DTSNRKNRNDLMKEFSYGT-----ASHSKVLGTYNQATVTVMEKEWKLIRI-- 300
 QY 272 LTFDDDRAGQAIAPKLNDEKGNADVEVLNLANHDKVFNANNLDGSPANIFEGEATSTTG 331
 Db 301 -----RVGAGFQW-----VDKNQNLNPKQE-----NLFEGK-AIIDPHGGIDSG 340
 QY 332 KLAVGIRKGGYKVEVQ-----VTKRGLTVSNITGIITVK-NLDTPASAIKNVFDALD 384
 Db 341 NVGYERESDVLDSRLKLIKAKAPFTVMTRTDNTPGNSDLSLKRVEFAQEHN 400
 QY 395 NDGVVYTGSKLGDPAFNQNL-----VVEKASLNKLVAITAGED 426
 Db 401 GDIFVS-HANGSAFKNGQGTETLYQSARAKVTNPNHVEDSKLLAQKIS-DRLVAAIGTKD 459
 QY 427 KVDPDGSIKSSNHGIISVVNNYVIAEAAAGEATLT 462
 Db 460 RGVKHQDL-----YVTEENTLPVLT 480

RESULT 6
 ABU17384
 ID ABU17384 standard; protein; 379 AA.
 XX

AC ABU17384;
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by Prokaryotic essential gene #2911.
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 OS *Bacillus anthracis*.
 PN WO200277183-A2.
 PD 03-OCT-2002.
 PF 21-MAR-2002; 2002WO-US009107.
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA21254.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 45308; 1766pp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 379 AA;

Query Match 10.9%; Score 459; DB 6; Length 379;
 Best Local Similarity 32.5%; Pred. No. 5.8e-19;
 Matches 124; Conservative 49; Mismatches 110; Indels 98; Gaps 13;

```

QY      5 FDFVPAGHWAEGSINYLVKGAITCKPDGTGYPTES:DRASAAVIFTKILNLPVDENAQP 64
          :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     24 FDFVEA--WADKSVTYLVDKVQLSGYPDGTGSSDDTLDRASAATIMTALGIHIIDLNAKP 81
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY     65 SFPKDAKNIWSSKYIAAVSKAGVWKGDGENPYPGKITDRASFASMLVSVAIULKDKVNGEL 124
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     82 SFDKSNHWGTMYMPFAAKASIKVEGKGIFNPFSKVTRRAAMATMLVNAYKLQNKTSTNG 141
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY    125 VTTFEDLLDHWGEEKANILINLGSIVFGGWPNKSVSRAEAAQFTALTDKKYGGKDNA 184
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db   142 QSKFEFKGHMGWGKIPTNLIGFISVGTDNQWQNKBITRAEAAQLTAKT----- 192
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY   185 QAUVTDVKYSBETPLTLTGCTGLDKLSADDVLBGDKAVAIEASPDGTSAAVVLGGKAEN 244
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db   193 ----MLOYSHSNPLE-----NKTIIDDPHGGED-----PG 219
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY   245 KDLTVKRNRQSFTVFKEVKKLAWEKLT-----FDDRRAGQAIFKLNDEKGNADV 296
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db   220 KD-TKGLPXXKTVLDTSLRLQKL-LEKHPTPTVLLNSXSDTRTHGSKSSL-QERG----- 272
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY   297 EYVLNLANHDVKPVANNLDGSPANTFEG-GEATSI-----TKGLA 334
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db   273 KPAKTKQGDIINGH-----ANAENGKKTEITYYXSSEKNTPHVEKKGPFFGKIQ 326
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY   335 V-----GIKQGDYK 344
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db   327 TRLVDALOTRDRGVKHXGLHV 347
          :|::||::||::||::||::||::||::||::||::||::||::||::||:

```

RESULT 7	
ABU18670	
ID	ABU18670 standard; protein; 408 AA.
XX	
XX	ABU18670;
XX	
XX	19-JUN-2003 (first entry)
XX	
DE	Protein encoded by Prokaryotic essential gene #4197.
DE	
XX	
XX	Antisense: prokaryotic essential gene; cell proliferation; drug design.
KW	

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pct_sequences

RESULT 8
AAR80530
ID AAR80530 standard; protein; 1252 AA.
XX AC AAR80530;

RESULT 8
AAR80530

XX	22-DEC-1995	(first entry)	
DT	XX	B. sphaericus SLP.	
XX	XX	Surface layer protein; SLP; fusion protein; vaccine; antigen;	
DE	XX	surface expression; epitope.	
KW	XX	Bacillus sphaericus.	
XX	XX	Key	
XX	XX	Location/Qualifiers	
FT	XX	Peptide	
FT	XX	1..30	
XX	XX	/label= Sig_peptide	
PN	XX	W09519371-A2.	
XX	XX	20-JUL-1995.	
PD	XX	13-JAN-1995;	95WO-EP000147.
XX	XX	14-JAN-1994;	94GB-0000650.
XX	XX	(SOLV) SOLVAY SA.	
PA	XX	Debiaere RY, Desomer J, Dhaese P;	
PI	XX	WPI; 1995-263827/34.	
DR	XX	N-PSDB; AAQ99430.	
XX	XX	Host cell expressing surface layer protein fusion protein - used for host	
PT	XX	presentation of antigens and vaccine prodn.	
XX	XX	Disclosure; Fig 6; 95pp; English.	
XX	XX	A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG P-	
CC	XX	13855); surface layer protein was used to screen an HindIII-generated	
CC	XX	library to isolate the slp gene. Promoter regions of the gene are used in	
CC	XX	genetic constructs providing surface expression of heterologous proteins	
XX	XX	in P-1 hosts	
XX	XX	Sequence 1252 AA;	
XX	XX	Query Match	8.8%; Score 369; DB 2; Length 1252;
XX	XX	Best Local Similarity	22.9%; Pred. No. 5.4e-13;
XX	XX	Matches	221; Conservative 131; Mismatches 347; Indels 264; Gaps 46
QY	13	WAEGSI NYLVDGAIATGTPGDTGPTESIDRASAAVFTKILNLPVDENAAQSFKDAAK-N 71	
DB	42	YAKEAVALVDGVLQGTNGNFPLNVTYRQAEEFTKALEL--EANGDYNFKDVRAG 99	
QY	72	IWSSKYIAAVERAGVVGDKGENFYPEGKIDRASFASMLYSAYNLKDKVNGELVTTFDL 131	
DB	100	AMTYNSIAAIVANGIFEGVSATFAPNKSILTRSEAAKILVEAFGLEGEAD---LSEFADA 156	
QY	132	--LDHWGEKEKANILNLGISVGT--GGKEPNKVSRAE--AAQFTALTDKKYGGKDNAAQAY 187	
DB	157	SQVKPWAKYKYLEIAVANGIFEGTDANKLPNNSTITQDFALVFRVTDKVEGSTPEEAA 216	
QY	188	V-----TDVKVS---EPTKLLITGTGLDKLSADDVTLEGDKAVAIEASTDGTSA----VV 235	
DB	217	VKAINNNTTVEVTFEEEVINQALNFKIEGLEIKNASVKQTNKKVVLVITTEAQSADKEYVL 276	
QY	236	TLGGK-----VAPNKDLTVKKVQKQSFVTFYVEVKKLAVEKLTPDDDBAGCAIAF 285	
DB	277	TLDTGETIGGFKGVAAVVPTK---VELVSSAVGQKLGQEVKQAKVTVAEGQSGAGIPVTF 333	
QY	286	-----KLNDEKGNADVEYLNLANHVDKFA-----NNLDGSPANIFEGG 324	
DB	334	TVPENNNDGWPTLTGTEALTNREGIATSYTRYKEGTDVTAATGDRSKFSILGYFWGY 393	
QY	325	EAPSTTGKLAVG--IKQG---DYKVEVQVTKRGLTVS---NTGIITVKNL---DTPASAI 374	
DB	394	DTLLSVEYVTTGASVNNNGANKTKYTVYKRNPKTGPEANKTFNTVGFVENNMVTSRDKVANAT 453	

PA (SLEY/) SLEYTR U.
 XX Lubitz W, Sleytr U, Kuen B;
 XX WPI; 1997-394558/37.
 DR N-PSDB; AAT75488.
 XX
 XX Preparation of S-layer proteins by expressing sbs-A gene in Gram negative
 PT bacterium - or new sbs-B gene in any host, also new recombinant proteins
 PT containing heterologous inserts, e.g. epitope(s), useful as vaccines and
 PT adjuvants.
 XX
 XX Claim 26; Page 19-23; 31pp; German.
 PS
 XX The present sequence is the Bacillus stearothermophilus PV72 S-layer
 CC protein, sbs-B. S-layer structures can be used as vaccines or adjuvants,
 CC particularly when they include a bacterial ghost that may contain
 CC additional epitopes in its membrane. Other uses of recombinant sbs-B,
 CC depending on the nature of the inserted peptide, are as an universal
 CC carrier for biotinylated reactants for use in immunological or
 CC hybridisation assays (the insert is streptavidin), to induce immune
 CC responses (epitopes), as a reagent for removing cytokine or toxin from
 CC serum (antigenic epitopes), as a molecular spinning nozzle
 CC (polyhydroxybutyrate synthase) and as a molecular laser (luciferase).
 CC (Updated on 17-OCT-2003 to standardise OS field)
 XX
 XX Sequence 921 AA;
 SQ
 Query Match 8.7%; Score 367.5; DB 2; Length 921;
 Best Local Similarity 24.9%; Pred. No. 4.4e-13;
 Matches 220; Conservative 121; Mismatches 335; Indels 209; Gaps 47;
 QY 4 SPFDVAGHWAEGSINLVLDGAIKPGDGYGPPESIDRASAAVIFTKILNLPVDENQA 63
 DB 33 SFTDVAPQI--KDAIDFVSTGATGKTETKFGVVDVETRLDAVILARVLEKLDVDAKD 90
 QY 64 PSFKDAKNIWSKYIAAWEKAGVWKGKGFYPEGKIDRASAFASMLVSAYNLKDKVNGE 123
 DB 91 AGFTDVPKD--RAKYNALVEAGLVNGKAPGFGAYDPLTRVEMAKIIANRYKLK--ADD 146
 QY 124 LVTTEDLLDHGE-----EKANILNLGISVGTGKWEKPNKVSRAAQAQIALTD 175
 DB 147 VKLPFTDVNDTWPVYKALYKVEVTRKLEKHOQASVHT-----KNITLRDPAQFVY--- 196
 QY 176 KXYGKDKNAQAVTDVQVSEPKLTLTGTGLDKLSADDVT--LEGDKAVAEAST---DGT 231
 DB 197 ----RAVNINAVPEIVEVAVNSITVKYTFNTQIADVDFTNFADNGLTVTKATLSRDKX 252
 QY 232 SAVVTLGGKAPNKLTVK---VKH-----OSFTVKFVYEVKKLAVKELTFDDDRAGQA 282
 DB 253 SVEVVVYKPFTRNQEVYITATGIRKLGKETAKELTGKFWVSQDVAVTVALANSSSLKVGE 312
 QY 283 IAFKLNDEKNADVEYLNLANHDVAFVANNLDGSPANI--PEGGEATSTTGKLVAGIKQG 340
 DB 313 SGLTVKQDQK-----DVGAKVELTSSNTNIVVYSSGEVSVSAKV-TAVKPG 360
 QY 341 DYKVEVQVTKRGGLTVSNTGIIITVKNLDTTPASAIKNVVFALDADNDGVVYVYKSLSGKDF 400
 DB 361 TADVTAKVTLPGVWLNTFKVTVTEVPV---QVQNGQFTL--VDN-----LSNA-- 405
 QY 401 ALNSQNLVVGKASLNKLAVATAGDKVY-----DPGS-----ISIKSSNHGIIIS 445
 DB 406 ----PQNTAFNKA--EKVTSMPAGETKTVAMVDYKNGDPEKFPDVKATVRSNLPIAT 460
 QY 446 VVNN-----YITAEA--AGBATITIKVDVTK-----DVK-----FKVTTDSRKLVSVK 487
 DB 461 AAINGSELLVATANAGQSGKASPEVTLKONTKRTFTVDVKDQVLDQIKVDATSVKLSDEA 520
 QY 488 ANPDKLVQVQNTKLPVTFVTTQYQDPPGANTAAKVEVLPKTVGVAEGGLDVTTDSSGI 547
 DB 521 VGGGEVGVNQKTIKVSAY--DOYG-----KEI--KFT--KGKVTVTINTGL 563
 QY 548 GYTKTIGVTDVGEQGVTFHPGNGATGLSLYVNVTEGNVAFKNFELSVKGYQGSPTK 607

Db 564 VIKNVN-----SDNTIDFSGNSAT-----DQFVVA-----TK 592
 QY 608 LDLNVSTTVVEYCLSKVTSRVVSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKV 667
 Db 593 -DKVNGKVEYFKNASD---TFTSTKITVNVNVK-ADATPVGLDIV---APSKI 643
 QY 668 DIHLTKNGATAGKATVEIVQ-ETIAIKSVN-----FKPVQTE---NFVEKKINIGTV 715
 Db 644 DVNAPN---TASTADVDFINFESVEIYTLDSNGRQKVKVTPATTTLVGTGKKKKVYV- 699
 QY 716 LELEKSNLDDIVKGINLTKETQHKVYVVKSGAE-----QGLKYLDRNGDAVFNAGDVKL 769
 Db 700 LQFKNE-----ELTLSTSSSTGNDGTAGMTKRIPIGK-YI--NSASVPASATVAT 748
 QY 770 GGVTVS-QTSDSALPNFKADLYDTLTITTKYDKGTLVFKVLKDKOV 813
 Db 749 SPTVKLNSSDNDL-----TEELIFGVIDPTQLV-----KDEDI 783

RESULT 10
 AAB10626
 ID AAB10626 standard; protein; 921 AA.
 XX
 AC AAB10626;
 XX
 DT 12-SEP-2003 (revised)
 DT 08-JAN-2001 (first entry)
 XX
 DE B. stearothermophilus sbsB protein.
 XX
 KW S-layer gene; sbsB; antibacterial; vaccine; adjuvant; bio reactor;
 KW polyhydroxyalkanoate synthesis.
 XX
 OS Geobacillus stearothermophilus.
 XX
 Key Location/Qualifiers
 FT Peptide 1..31
 FT /label= signal_peptide
 FT Peptide 32..921
 FT /label= mature_peptide
 XX
 PN DE19903345-A1.
 XX
 PD 03-AUG-2000.
 XX
 XX 28-JAN-1999; 99DE-01003345.
 PF
 XX 28-JAN-1999; 99DE-01003345.
 PR
 XX
 XX (LUBI/) LUBITZ W.
 PA
 XX Lubitz W;
 PI
 XX WPI; 2000-533868/49.
 DR N-PSDB; AAA71798.
 XX
 XX Host cell, useful e.g. as bioreactor for production of
 PT poly(hydroxyalkanoate), containing two or more recombinant polypeptides,
 PT with at least one in carrier-bound form.
 XX
 XX Disclosure; Page 20-22; 26pp; German.
 PS
 XX This invention describes a novel host cell (A) comprising at least two
 CC functional recombinant polypeptides (1), at least one being in carrier
 CC bound form. The products of the invention have antibacterial activity.
 CC (A), or, where bacterial, their ghosts (B), are useful as vaccines or
 CC adjuvants (specifically for presentation of immunogenic epitopes of
 CC pathogens or autologous immunostimulatory polypeptides, e.g. cytokines),
 CC or preferably, as enzyme reactors for performing a cascade of reactions,
 CC specifically synthesis of poly(hydroxyalkanoate). Localization of
 CC individual (1), specifically enzymes, in separate cellular compartments
 CC avoids adverse reactions between products and substrates, when being used

Db 1483 GLNIISNGRNTVRLRGKIDVKYIQGVASVEEVIEAKRVLEKVKDLSDERETAKLG 1542
 QY 730 IN-----LTKETQHKVRVVKSG---AQQKLYLDRNGDAVFNAGD 766
 Db 1543 VSAVRVFEVPEPNAITVNTQNEFTTKPSSQVITSECK-----ACFSSGN 1584
 RESULT 14
 ID AAR77395 standard; protein; 941 AA.
 AC AAR77395;
 XX
 DT 10-APR-1996 (first entry)
 XX
 D2 Full length Bacillus sp. alkaline cellulase.
 XX
 KW Alkaline cellulase; Bacillus; E.coli; pBR322; expression vector; stable;
 KW B.subtilis; specific activity; surface active agent; chelating agent.
 XX
 OS Bacillus sp.
 XX JP07203960-A.
 PN
 XX 08-AUG-1995.
 XX
 XX 24-JAN-1994; 94JP-00005714.
 XX
 PR 24-JAN-1994; 94JP-00005714.
 XX
 PA (KAOS) KAO CORP.
 XX
 DR WPI; 1995-307162/40.
 DR N-PSDB; AAQ94350.
 XX
 PT Alkaline cellulase and related DNA, vectors and transformed microbes -
 PT useful for prodn. of a highly specific enzyme in alkaline environments.
 XX
 PS Disclosure; Page 10-14; 17pp; Japanese.
 XX
 CC The amino acid sequence of the full length alkaline cellulase. The gene
 CC encoding the protein was isolated from Bacillus sp. KSM-365 chromosomal
 CC DNA and inserted into the E.coli vector pBR322. The gene was subsequently
 CC used to produce expression vectors contg. a 1071 fragment of the gene
 CC covering amino acid residues 228-584 (AAQ94349). The expression vectors
 CC contg. the fragment were transformed into B.subtilis for production of
 CC the protein. The novel alkaline cellulase has high specific activity in
 CC an alkaline environment and is stable over wide ranges of temp. and pH,
 CC and has full activity in the presence of surface active agents and
 CC chelating agents
 XX
 SQ Sequence 941 AA;
 Query Match 6.0%; Score 252.5; DB 2; Length 941;
 Best Local Similarity 21.3%; Pred. NO. 2.6e-06;
 Matches 193; Conservative 109; Mismatches 291; Indels 311; Gaps 42;
 QY 5 PDVPAQHWAGSINLYDKGAIKPGDTGPTESIDRASAIVFTKILNLPVDENAQP 64
 Db 41 FSDVKTSWSPFYIKDLVEQEVITGTSATTSPTSDSVTRAQFTVMLRGLGLEASSKDYP 100
 QY 65 SPKDAKNWSSKYIAAVEKAGVVGKDGKGFENFYPEGKIDRASFAFMLVSAYNL----- 116
 Db 101 -FKDRKN-WAYKEIOAAVEAGIVTGKTNGEFAPNENITREQMAAMAVRAYEYLENLSLP 158
 QY 117 ---KQKANGELVTTT-EDLDHWGKEKAILNIGISVGTGKWEPNKSVRAEAAQFIA 172
 Db 159 EQREYNDSSTSTTAQDAV-----QKAYVLELM--EGNTDGYQPKENSTREQSAKVIS 211
 QY 173 LFDKXGKONKQAQVTDVKVSEPK-----LTLTGGLDKLSADDVTLEGDKAVAE- 225
 Db 212 TLLWKVASHD--YLVHTEA-VKSPSEAGALQIVELNG-----QUTLAGEDGTVPQL 259

QY 226 --ASTDGT---SAVVTLGGKVAPNKKD-----LTVKVKQNSFVTK-----FVYEVKKLA 268
 Db 260 RCMSTHGLQWFEIENVENAFVALSDNGSNIRLAMYIGENGATNPEVKDLVYEGIELA 319
 QY 269 VEKLTFF-----DDDRAG-----QATAFKLNDEKGNADVEYL--NLANHDKV 307
 Db 320 FEHDMYIVDVHWHAPGDPRADYVSGAYDFEIEADHYKHDPKQ---HYITWELANEP-- 374
 QY 308 FVANLNDGSP--ANIFEGGEATSTTKLAVGI--KQGDYKVEV---QVTKRGGTLVSNVTG 360
 Db 375 --SPNNGGPGLTNDEKWEAVEPIVEMLEKKGDMMLVGNPNWSQFPLSADNP- 431
 QY 361 IITVKNL-----DTPAS-----AIKNVVFALD----- 382
 Db 432 -IDAENIMYSVHFYTGSHGASHIGYPECTSPSSRSNNVNAVRYALDNGVAVFATEWGTQS 490
 QY 383 ADNDG-----VVNYGSK--LSGKDFALNSONLVVG--EKASLNKLVAIAGEDKVV 429
 Db 491 ANGDCGPFDEADWLNFLNKHNSWANSLTNKNEISGAFTPELGRTDAT-----DL 544
 QY 430 DFGSISIKSNHGIISVVNNYITAEAGEATLTIKVGDVTKDVK-----FKVT 477
 Db 545 DFGANQVWAPEE--LSLSGEYVRARIKGIETPI---DRTKFTKLAVMDFNDGTTQGFQVN 599
 QY 478 TDS--RKLVSVKANPDKLQV---VQN----- 498
 Db 600 GDSPNKRESITLNNNDALQIEGLNVNSDISGNVWDNVRLSADGWSNVDILGATELTID 659
 QY 499 -----KTLPTVFTTQYQDPPFGANTAAIKVLPKTVGV 532
 Db 660 VIVEBPTTVSIAAIPQGAAGWANPTRAIKVTEDDFESFGDGYKALVITISDPSLETI 719
 QY 533 AEGGLD-----VVTDSGSGICTKTIQVTDNDVGGSTVHFONGGATLGSLY---- 578
 Db 720 ATSPEDNTWSNILLFVGTEADAVISLDNITVSGTIEBIEVHDEKGT-ATLPSTFEDGTR 778
 QY 579 -----VNVTEGN-----VAFKMFELVKVQG 599
 Db 779 QGMDWHTESGVKTAITIEANGSNALSWEYAVEKPSDGNATAPRLDFWKDELVRGTSD 838
 QY 600 Y-----GSPDTKLDLNVSTTVYQLSKYVTSDRVYSDPENLEG 637
 Db 839 YISFFPYIDAVRASEGALSINAVFOPPANGYQWVEPTTFEIDLTELDSDATVTS--ELYH 896
 QY 638 YEVE 641
 Db 897 YEVK 900
 Cellulase.
 Enzyme.
 Bacillus sp; KSM-634.
 Key Location/Qualifiers
 Protein 1..584
 /note= "claim 2"
 JP01281090-A.
 13-NOV-1989.

PF 02-MAY-1988; 88JP-00109545.
XX PR 02-MAY-1988; 88JP-00109545.
XX PA (KAOS) KAO CORP.
XX DR WPI; 1990-330487/44.
XX DR N-PSDB; AAO06320.
XX PT DNA segment contg. cellulase gene - originated from alkalophilic Bacillus
XX PT and codes cellulase.
XX PS Claim 1; Fig 1; 21pp; Japanese.
XX CC DNA encoding the sequence can be introduced into E.coli to give a
XX CC recombinant strain for prodn. of cellulase protein. (Updated on 27-AUG-
XX CC 2003 to correct OS field.) (Updated on 24-OCT-2003 to standardise OS
XX CC field)
XX SQ Sequence 941 AA;
Query Match 6.0%; Score 251.5; DB 2; Length 941;
Best Local Similarity 21.3%; Pred. No. 2.9e-06;
Matches 193; Conservative 108; Mismatches 292; Indels 311; Gaps 42;
Qy 5 FPOVPAGHWAEGSINLYVDKGAITKPGDGTGPTESIDRASAIVITKILNLPVDENAOP 64
Db 41 FSDVKTSWSFPYIKDLYEQEVITGATTFSTDSVTRAQFTVMTLRLGLEASSKDYP 100
Qy 65 SFKDAKNWSSKXIAAWEKAGVVKGDGKGFYPEGKIDRASFASMLVSAYNL----- 116
Db 101 -FKDRKN-WAYKEIQAAEYAGIVTGTNGEFAFENITREQMAAMAVRAYEYLENELSLP 158
Qy 117 ---KDKNGELVTFP-EDLLDHGEEKXANILNLIGISVGTGGKWEKNKSVSRAEAQAFTA 172
Db 159 EEOREYNDSSISFQAQDAV-----QKAYVLELM--EGNTDGYFPQKRNSTREQSAKVIS 211
Qy 173 LTKKYGKDKNAQYVTDVKVSEPTK-----LTLTGTGLDKLSADDVTELEGDKVAIE- 225
Db 212 TLLWKVASHD--YLYHTEA-VKSPSEAGALQLVELNG-----QLTLAGEDGTPVOL 259
Qy 226 --ASTDGT---SAVVTGGKVAPKND-----LTVKVKQNSFVTK-----FVYEVKKLA 268
Db 260 RGMSTHGLQWFEIVNENAFVLSNDWGSNMIRLAMYIGENGYATNPVKDLVYEGIELA 319
Qy 269 VEKLT-----DDDRAG-----QAIAFKLNDEKGNADVEYL--NLANHVK 307
Db 320 PEHDMYIVDMWHVAPGDPRAVDVSGAYDFFEEIADHYKHDPKN---HYIIWELANEP-- 374
Qy 308 FVANNLDGSP--ANIFEGGEATSTGKLVGI--KQDGYKVEV---QVTKRGGLTVSNTG 360
Db 375 --SPNNGGPGLTNDEKGEAVKEYAEPIVEMLEKRGKDNMILVGNPNWSQRPDLSADNP- 431
Qy 361 IITVKNL-----DTPAS-----AIKNVVPALD----- 382
Db 432 -IDAENIMYSVHFYFTGSHGASHIGYPEGTPSSERSNVMANVLLDNGVAVFATWGTQS 490
Qy 383 ADNDG-----VWYVYGSK--LSGKDFALNSQLVVG--EKASLNKLVIATIGEDKV 429
Db 491 ANGDDGPFYDEADVNLNFKHNISWANWSLTNKNESIGAFTPPELGRDTAT-----DL 544
Qy 430 DPGSISIKSSNHGIIISVNNVITAEAGEATLTIKVGDVTKVK-----FKVT 477
Db 545 DPGANQVWAPEE--LSLSGEYVRARIKGIETPI---DRTKFTKLWVDFNDGTQGFQVN 599
Qy 478 TDS--RKLVSVKANPKLQV-----VQN----- 498
Db 600 GDSPNKESITILSNNDALQIEGLNVNDISBGNVWNVRLSADGWSENVLDIGATELTID 659
Qy 499 -----KTLPVTFVTTDQYGDPPGANTAAIKVLPKTCVV 532
Db 660 VIVEEPTTVIAIPQGPAGANPNRAIKVTEDDFESFGDGYKALVTTITSEDSPSLETI 719

Qy 533 AEGGLD-----VVTDSGSGTGTITGVTDNDVGEGTVHFQNGNGATILGSLY----- 578
Db 720 ATPEDNTMSNIILFVGTEADVISLDNITVSGTEIEIEVIHDEKGT-ATLPSTFEDGTR 778
Qy 579 -----VNVTEGN-----VAFKNFELVSKVGQ 599
Db 779 QCNWHTTESGVKTAITTEEANGSNALSWEYAYPEVKPSDGMATAPRLDFWKDELVRGTSD 838
Qy 600 Y-----CQSPDTKLDLNVSTTVVEYQLSKYTSDRVYSDPENLEG 637
Db 839 YISFDIFYIDAVRASEGAISINAVFPFPANGYQWQVPTTFEIDLTELDSATVTS-ELYH 896
Qy 638 YEVE 641
Db 897 YEVE 900
Search completed: April 7, 2004, 17:26:34
Job time : 68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:23:54 ; Search time 46 Seconds
(without alignments)
4755.947 Million cell updates/sec

Title: US-09-844-281-1
Perfect score: 4202
Sequence: 1 AGKSPDPVAGHWAGSINY.....ITSETGQAVHVNLPNL 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 262633353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4202	100.0	833	9	US-09-844-281-1
2	564.5	13.4	529	12	US-10-282-122A-45596
3	556.5	13.2	414	12	US-10-282-122A-45560, A
4	533.5	13.2	525	12	US-10-282-122A-46674, A
5	459	10.9	379	12	US-10-282-122A-45308
6	374	8.9	1222	9	US-09-137-531-15
7	374	8.9	1252	9	US-09-137-531-9
8	369	8.8	408	12	US-10-282-122A-46594, A
9	367.5	8.7	921	9	US-09-117-447-6
10	257.5	6.1	1599	13	US-10-092-880-9
11	249.5	5.9	1600	13	US-10-092-880-10
12	249.5	5.9	2353	9	US-09-797-862-33
13	246	5.9	1180	14	US-10-193-764-61
14	246	5.9	1188	14	US-10-193-764-59
15	242.5	5.8	1220	14	US-10-193-764-28

16	242.5	5.8	1226	14	US-10-193-764-26	Sequence 26, Appl
17	235	5.6	1004	14	US-10-193-764-53	Sequence 53, Appl
18	235	5.6	1010	14	US-10-193-764-51	Sequence 51, Appl
19	235	5.6	2712	12	US-10-282-122A-67070	Sequence 67070, A
20	230.5	5.5	2154	12	US-10-282-122A-66695	Sequence 66695, A
21	230	5.5	1325	9	US-09-741-669-304	Sequence 304, Appl
22	229	5.4	969	14	US-10-193-764-32	Sequence 32, Appl
23	229	5.4	975	14	US-10-193-764-30	Sequence 30, Appl
24	227.5	5.4	1073	14	US-10-193-764-45	Sequence 45, Appl
25	227.5	5.4	1079	14	US-10-193-764-43	Sequence 43, Appl
26	223.5	5.3	852	12	US-10-282-122A-43144	Sequence 43144, A
27	222	5.3	1833	14	US-10-175-282-4	Sequence 4, Appl
28	222	5.3	1833	14	US-10-175-275-4	Sequence 4, Appl
29	222	5.3	1992	14	US-10-175-282-3	Sequence 3, Appl
30	222	5.3	1992	14	US-10-175-275-3	Sequence 3, Appl
31	222	5.3	2265	12	US-10-282-122A-45123	Sequence 45123, A
32	220.5	5.2	4327	15	US-10-369-493-10178	Sequence 10178, A
33	218.5	5.2	2122	9	US-09-813-214A-9	Sequence 9, Appl
34	217.5	5.2	1005	14	US-10-193-764-41	Sequence 41, Appl
35	217.5	5.2	1011	14	US-10-193-764-39	Sequence 39, Appl
36	217	5.2	1483	12	US-10-282-122A-51483	Sequence 51483, A
37	216.5	5.2	1095	14	US-10-193-764-65	Sequence 65, Appl
38	216.5	5.2	1536	13	US-10-092-880-2	Sequence 2, Appl
39	216.5	5.2	1536	14	US-10-193-764-63	Sequence 63, Appl
40	216	5.1	3217	12	US-10-282-122A-61210	Sequence 61210, A
41	215.5	5.1	1222	14	US-10-193-764-37	Sequence 37, Appl
42	215.5	5.1	1228	14	US-10-193-764-34	Sequence 34, Appl
43	214.5	5.1	1946	12	US-10-282-122A-62947	Sequence 62947, A
44	213	5.1	1098	9	US-09-797-862-32	Sequence 32, Appl
45	211.5	5.0	3241	9	US-09-841-786-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

```
US-09-844-281-1
; Sequence 1, Application US/09844281
; Patent No. US20020082386A1
; GENERAL INFORMATION:
; APPLICANT: Mangold, Beverly L.
; APPLICANT: Aldrich, Jennifer L.
; APPLICANT: O'Brien, Thomas
; TITLE OF INVENTION: Anthrax Specific Antibodies
; FILE REFERENCE: 38602.0003
; CURRENT APPLICATION NUMBER: US/09/844,281
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/200,505
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-844-281-1
```

Query Match 100.0%; Score 4202; DB 9; Length 833;
Best Local Similarity 100.0%; Pred. No. 1.1e-290;
Matches 833; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AKKSPDPVAGHWAGSINYLVDKGALTGKPDGYGPTESIDRASA	AVFTKILNLPYDE	60
DB	1	AKKSPDPVAGHWAGSINYLVDKGALTGKPDGYGPTESIDRASA	AVFTKILNLPYDE	60
QY	61	NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKGFPEGKIDRASFASMLVSNLKDVK	120	
DB	61	NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKGFPEGKIDRASFASMLVSNLKDVK	120	
QY	121	NGELVTTTFEDLLDHWGEEKANILNLGISVGTGKWEKNSVRAEAAQFTALTDKKGK	180	
DB	121	NGELVTTTFEDLLDHWGEEKANILNLGISVGTGKWEKNSVRAEAAQFTALTDKKGK	180	

QY 181 KDNAQYVTDVKSPTKLTGTGLDKLSADVTLEGDKAVAEASTDGTSAVVTLLGGK 240
 Db 181 KDNAQYVTDVKSPTKLTGTGLDKLSADVTLEGDKAVAEASTDGTSAVVTLLGGK 240
 QY 241 VAPNKLTVKVNQSFVTFYVEVKLAVEKLTEDDDRAGQAIAFKLNDEKGNADVLYN 300
 Db 241 VAPNKLTVKVNQSFVTFYVEVKLAVEKLTEDDDRAGQAIAFKLNDEKGNADVLYN 300
 QY 301 LANHDKVFNANLDSGANIPFEGGEATSTTGKLAIVGKQGDYKVEVQVTKRGGTLVNTG 360
 Db 301 LANHDKVFNANLDSGANIPFEGGEATSTTGKLAIVGKQGDYKVEVQVTKRGGTLVNTG 360
 QY 361 IITVKNLTPASAIKNVFPALDADNDGVNVTGSKLSDGDFALNSQNLVVGESKASLNKLA 420
 Db 361 IITVKNLTPASAIKNVFPALDADNDGVNVTGSKLSDGDFALNSQNLVVGESKASLNKLA 420
 QY 421 TIAGDKVVDPSGISIKSSNHGIIISVNNYITAEAGAEATLTIKVGDVTKDVKFKVTTDS 480
 Db 421 TIAGDKVVDPSGISIKSSNHGIIISVNNYITAEAGAEATLTIKVGDVTKDVKFKVTTDS 480
 QY 481 RKLVSVKANPKLQVQNKTLPTVTFVTTDQYGDPPFGANTAAIKVLPKTVGAEGGLDVV 540
 Db 481 RKLVSVKANPKLQVQNKTLPTVTFVTTDQYGDPPFGANTAAIKVLPKTVGAEGGLDVV 540
 QY 541 TTDSSIGTKLTGVNGDVGEQTHFONGATLGSLYVNTGNAVAFKPFELVSKVGOY 600
 Db 541 TTDSSIGTKLTGVNGDVGEQTHFONGATLGSLYVNTGNAVAFKPFELVSKVGOY 600
 QY 601 GOSPTKLDLNVSTTVYQLSKYTSRDVYSDPENLEGEVESKNLAVADAKIVGNKVVT 660
 Db 601 GOSPTKLDLNVSTTVYQLSKYTSRDVYSDPENLEGEVESKNLAVADAKIVGNKVVT 660
 QY 661 GKTGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVOTENFVEKINIGTVLEBK 720
 Db 661 GKTGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVOTENFVEKINIGTVLEBK 720
 QY 721 SNLDDIVKGINLTKEQHKVRVWKSAGQGLYLDNRGDAVFNAGDVKLGDTVVSQTSDS 780
 Db 721 SNLDDIVKGINLTKEQHKVRVWKSAGQGLYLDNRGDAVFNAGDVKLGDTVVSQTSDS 780
 QY 781 ALPNFKADLYTLTKYTKDGLTKVTKVTKDQVITSEIGSOAVHVNVLNPNL 833
 Db 781 ALPNFKADLYTLTKYTKDGLTKVTKVTKDQVITSEIGSOAVHVNVLNPNL 833

RESULT 2

US-10-282-122A-45596
 ; Sequence 45596, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Cart, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 45596
 ; LENGTH: 529
 ; TYPE: PRI
 ; ORGANISM: Bacillus anthracis
 ; US-10-282-122A-45596
 Query Match 13.4%; Score 564.5; DB 12; Length 529;
 Best Local Similarity 32.6%; Pred No 1,1e-31;
 Matches 172; Conservative 81; Mismatches 192; Indels 83; Gaps 18;
 QY 1 AKSPDPVAGHWAGSINYLVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDE 60
 Db 25 AAKTFSDVPT--WAQSDVYLVGKALDKGPDGTSPSEAVDRGSAAKILAVLGLPVDP 82
 QY 61 NAQSPFKDAKNTWSSKYIAAVEKAGVWKGDKGKENVPEKIDIRASFAFMLVSAYNLKDKV 120
 Db 83 KAKSPFKDAKNTWAPYIAAVEKAGVINGDGTGKFNPSQINRASVAMLVQAYSLDKKI 142
 QY 121 NGBELVTFPDLDDHGEEXANILINLGHISVGTGGKWEPNKSVSRAEAAQFIALT- 175
 Db 143 IGEPTQFKDLBPHWGKQANILVALET-SKGTGNGWNEGTVTTRAEEAQFIAMADQNKTS 202
 QY 176 --KYYGKKDNAQAY-----VTDVKVSEPTKLTGTGLDKLSADVTLEGDKRAVLE 225
 Db 203 TSKRMVMNRVITYHQPSLSGSDTDVQ-HKPMQVEVT-----E 239
 QY 226 ASTDGTSAVVT-LGGKVAENKDLTVKVK--NOSFVTKFYVEVKLAVEKL-TFDD- 276
 Db 240 QRADGWLKIVTSKGEKWP--LTEKTETINEEFTT--YETASHSSKVLGTYNACTVTV 293
 QY 277 -DRAGQAIAFKLNDEKGNADVLYNLANHDKVFNANLDSGANIPFEGGEATSTTGKLA 335
 Db 294 MEESGSWIRIVRGAGFQWYDKNQLNPVKQB-----NFLEGG-AIITDPGHGIDSQNGVY 347
 QY 336 GIKQGDYKVEVQV-----TKRGLTVSNTGIITVK-NLDTPASAIKNVFPALDADNDGV 388
 Db 348 YEKESETVLDVSLRLKKIPEQKAPFTVMFTRTDTRPGVNSTDSLKKRVEFAQEHNGDIF 407
 QY 389 VNYGSKLSGKDFALNSQNL-----VVGEKASLNKLVATTAGDKVVDPSGISIKSSN 440
 Db 408 VSHANGSAEKNGQGTETILYQSAKAVTNPHVEDSKLLA-----QKIQDLVLAALGTDK 462
 QY 441 HGIISVNNYITAEAGAEATLTIKVGDVTKDVKFKVTTDSRKLVSVKA 488
 Db 463 RG-VKQDLYVTRENTMPAULTAFVNDKSDAKIATPKQQAAREA 509
 RESULT 3
 US-10-282-122A-45560
 ; Sequence 45560, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari

QY 121 NGELVTTFFDLHDHGWGEEKANILINLIGISVGTGGKWEPEKNSVSRASAAQFIALTDKYYK 180
Db 138 IGEFTQFQDLPHGKQKQANILVALEISKGTGNGWNEPVGTVTRAEAAQFIAMAEKKIKQ 197
QY 181 KDNAQAY-----VTDV-----KVSEPTKLTGTGLDKLSA----- 211
Db 198 STSKRMVTRNVITVHHPSLSGGITVDVHKFPQNGKVTQ-----RADGWVKMLTSKREK 252
QY 212 DDTLTLEGDKAVAIEASTDGTSAVVTLGKGVAPNLDLTVKVNQSFVTKFVVEVKLAVER 271
Db 253 DTSNKNRNDLMKEFSTVGT-----ASHSKVLGTNAQTIVTMEKEWLIIRI-- 300
QY 272 LTFDDDRAGQAIAFKLNDKGNADVEYINLANHDVFKVANNLDGSPANIFEGGSAITSTG 331
Db 301 -----RVGAGFQW-----VDKNQLNPKQE-----NFELEGK-AIIIDPHGGGIDSG 340
QY 332 KLAVGIKQGDYKVEYQ-----VTKRGGLTVSNTGIIIVK-NLDTPASAIQNVVFFALDAD 384
Db 341 NVGYEYESDVTLDVSLRLKLIIXAKAFTVMTFRDTRPGVNSTSLKCRVEFAQEHN 400
QY 385 NDGVVNYGSKLGGDFALNSQNL-----VVGEEKASLNKLVIATIGED 426
Db 401 GDIFVSIHANGSAERKNGOGTETLYQSARAKVTNPNHVEDSKLLAQKIS-DRLVAALGTD 459
QY 427 KVVDPGSISIKSNHGIISVNNVITAEAGEATILT 462
Db 460 RGVKHQDL-----VYTRENTLPVAVLT 480

RESULT 5

US-10-282-122A-45308
; Sequence 45308, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Maloney, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45308
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Bacillus anthracis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (227)..(227)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (229)..(229)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (255)..(255)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (304)..(304)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (369)..(369)
; OTHER INFORMATION: X=any amino acid
; US-10-282-122A-45308

Query Match 10.9%; Score 459; DB 12; Length 379;

Best Local Similarity 32.5%; Pred. No. 2.2e-24;

Matches 124; Conservative 49; Mismatches 110; Indels 98; Gaps 13;

QY 5 FPDVPAGHWAEGSINVLDKGAITGKPDGTGPTESIDRASAAVFTKILNLPVDENAP 64
Db 24 FPDVPA--WADKSVTVLDKQVLSGYPDGTFGSSDTLDRASATMTTALGHIHDLNKP 81
QY 65 SPKDAKNIWSSKYIAAVEKAGVVGDKGXENFVPEKIDRASPSMLVSAYNLKDKVNGEL 124
Db 82 SPKDSQNHMGTPYMPAAEKAGSIKVEKGIFNPSSKVTAAAMATMLVNAYKLQNKNTSG 141
QY 125 VTFEEDLLDHGGEERANTILINLIGISVGTGGKWEPEKNSVSRASAAQFIALTDKYYKQDNA 184
Db 142 QSKFEFGHGGKIPNLTIGFEISVGTONGQPNKFTIRAEAAQLTAKTD----- 192
QY 185 QAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLLEGSKAVAIEASTDGTSAVVTLGKGVAPN 244
Db 193 ---MLQYSHSNPLE-----NKTIIIDPGHGDE-----PG 219
QY 245 KDLTVKVNQSFVTKFVVEVKLAVERKLT-----FDDDRAGQAIAFKLNDKGNADV 296
Db 220 KD-TKGLPKXIXIVLDTLSRLQKL-LEKHTPFTVLNKSXSDTRTGHGSKSL-QERG---- 272
QY 297 EYLNLANHDVFKVANNLDGSPANIFEG-GEATST-----TGKLA 334
Db 273 KEAKTKQGDILIMGH-----ANAFNGNGKRTETVYYXSSKSEKTNPHVKKFPFGKIQ 326
QY 335 V-----GIKQGDYKV 344
Db 327 TRLVDALQTRDRGVKKGDLHV 347

RESULT 6

US-09-137-531-15
; Sequence 15, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-137-531-15

Query Match      8.9%; Score 374; DB 9; Length 1222;
Best Local Similarity 23.3%; Pred. No. 1.5e-17;
Matches 224; Conservative 130; Mismatches 345; Indels 264; Gaps 47;

QY 13 WAEGSYNLVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDENAQPSFKDAK-N 71
DB 12 YAKEAVQALVDQGVIGDGTNGNPNLNTVTRAQAEIFTKALEL--EANGDVNFKDKVAG 69
QY 72 IWSSKYLIAAVEKAGVVGKDGKENFYPEGKIDRASPAFMLVSNLKDKNVNGELVTTFEDL 131
DB 70 AWYNSIAAVVANGIFEGVSATFAPNKSILTRSEAAKILVEAFGLEGEAD---LSEFADA 126
QY 132 --LDHWGEKANIILNLGISVGT-GGKWEPNKSVSRAB-AAQFIATLDKKYKKDNQAAY 187
DB 127 SQVKPWAKYLEIAVANGIFEGTDANKLPNNISITRQDFALVFKRTVDKVEGETPEEAAP 186
QY 188 V-----TDVKVS---EPTKLTITGGLDKLSADDVTL--EGDKAVAI---EASTDGTSAVV 235
DB 187 VKAINNTTVEVTFEEBVTNVQALNFKIEGLEIKNASVKQTKNKVVLTTEAQTADKEYVL 246
QY 236 TLGGK-----VAPNKDLTVKVNQSFVTKFVVEVKLAVEKLTFFDDDRAGQALAP 285
DB 247 TLDGETIGGFKGVAAVVPTK---VELVSAVQGLQGEVVKQAKVTVAEGQSKAGIPVTF 303
QY 286 -----KLNDEKGNADVEYLNLANHDKVFA-----NNLDGSPANIPEGG 324
DB 304 TVPGNNNDGVVPTLTGEALTNBEGIATYSYTRYKEGTDVATAYATGDRSKFSLGVVFWGV 363
QY 325 EATSTTGKLVG--IKQG---DYKVEVQVTKRGGITVS---NTGIITVKNL--DTPASAI 374
DB 364 DTLVSEEVTTGASVNNGANKYKYTKYKPKTGPEANKTNPVGVNNVTSOKVANAT 423
QY 375 KHWFPALDADNDGVVNYGSKLGGKDFALNSQNLVVGKASLNKLVATTAGEDKVVDPGSI 434
DB 424 VNGVKALQLSN-----GTALDAAQITTDSK-----GEATFTVSGTNAAVTPVYV 467
QY 435 SIKSSNHGIIISVNNYITAE-----AAGEATLTIKV-----GDV-----TKDV 472
DB 468 DLHSTNN---STSNKKYSASALQTTASKVTFAALQAEYTIETLRADNAGEVAATGATNGR 524
QY 473 KFKVTTDSR-----KLVSVKANPKDLQVQNKLTPLVTFTVTDQYGPFGANTAAIKEV 525
DB 525 EYKVIKQKAGNLAKNEIWNVAFNEDKDRVISTVT--NAKFVDTD-----567
QY 526 LPKTTGVVAEGG---LDVVTTDSGS-----IGTKT-----IGVTGNDVGGTV---564
DB 568 -EDTAVYETGDKRAKQISVKTNDKGEATFVIGSDTVNDYATPIAWIDINTSDAKQGDLEG 626
QY 565 -----HFO-----NNGNG-----ATLGLSYNNVTGNAFKNFKELVSKVGQ-----599
DB 627 EPKAVAPISYFQAPVLDGSAIKAYKSLNKAIVTKFDGSETAVFAAELVNGSKKVTGTS 686
QY 600 -----YQG--SPTKLDLNV-----STTVEYQLSKYTSDRVSDPENLEGYRVEKSNL 645
DB 687 IKKATVTIYNTGANDIKVDNQVISPNSRVTVTYEATLSSTGTGVTTPAKNLEVTISVDGKTT 746
QY 646 AYADAKIVGNKVVVVG---KTPGKVDIHLTKNGTAG--KATVEIVQETIATKSVNFKVQ 701
DB 701
```

```

Db 747 AV-----KVIAIGIAVNTDGK-DYAFTAKEATATFATNEVPN-----SYTGVA 789
QY 702 TE-NFVEKKINIGTVLEKSNLDDIVKGINLTQTKQHKVRVVKSGABOGKUL--YLDNRG 758
DB 790 TQFNTADSGSNSNSIWFAGKN-----PVKYAGVSGKTYKYFGANG 829
QY 759 DAVFNAGDVKLGDVTVTSQSDSALPNFKADLYDTLTITTKYTDKG---TLVFKVLKDKDVIT 815
DB 830 NEVFGEA-----ANEALLTQYATEGQKVTISYNV--DQDITV 864
QY 816 SEI 818
DB 865 FKV 867

RESULT 7
US-09-137-531-9
; Sequence 9, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-137-531-9

Query Match      8.9%; Score 374; DB 9; Length 1252;
Best Local Similarity 23.3%; Pred. No. 1.6e-17;
Matches 224; Conservative 130; Mismatches 345; Indels 264; Gaps 47;

QY 13 WAEGSYNLVDKGAITGKPDGTGPTESIDRASAIVFTKILNLPVDENAQPSFKDAK-N 71
DB 42 YAKEAVQALVDQGVIGDGTNGNPNLNTVTRAQAEIFTKALEL--EANGDVNFKDKVAG 99
QY 72 IWSSKYLIAAVEKAGVVGKDGKENFYPEGKIDRASPAFMLVSNLKDKNVNGELVTTFEDL 131
DB 100 AWYNSIAAVVANGIFEGVSATFAPNKSILTRSEAAKILVEAFGLEGEAD---LSEFADA 156
QY 132 --LDHWGEKANIILNLGISVGT-GGKWEPNKSVSRAB-AAQFIATLDKKYKKDNQAAY 187
DB 157 SQVKPWAKYLEIAVANGIFEGTDANKLPNNISITRQDFALVFKRTVDKVEGETPEEAAP 216
QY 188 V-----TDVKVS---EPTKLTITGGLDKLSADDVTL--EGDKAVAI---EASTDGTSAVV 235
DB 217 VKAINNTTVEVTFEEBVTNVQALNFKIEGLEIKNASVKQTKNKVVLTTEAQTADKEYVL 276
QY 236 TLGGK-----VAPNKDLTVKVNQSFVTKFVVEVKLAVEKLTFFDDDRAGQALAP 285
DB 277 TLDGETIGGFKGVAAVVPTK---VELVSAVQGLQGEVVKQAKVTVAEGQSKAGIPVTF 333
QY 286 -----KLNDEKGNADVEYLNLANHDKVFA-----NNLDGSPANIPEGG 324
DB 334 TVPGNNNDGVVPTLTGEALTNBEGIATYSYTRYKEGTDVATAYATGDRSKFSLGVVFWGV 393
QY 325 EATSTTGKLVG--IKQG---DYKVEVQVTKRGGITVS---NTGIITVKNL--DTPASAI 374
DB 374
```

```
Db 394 DTILSVEVTTGASVNNGANKTYKYTKNPKTGKPEANKTNFVFNENNTSDKANAT 453
Qy 375 KNVFALDADNDGVVNYGKLSGKDPFALNSQNLVVGKASLNKGLVATTAGEDKVVDPGSI 434
Db 454 VNGVKALQLSN-----GTALDAAQITTDK-----GEATFTVSGTNAATPVVY 497
Qy 435 STKSNHGIISVNNYITAE-----AAGEATLTIV-----GDV-----TKDV 472
Db 498 DHSTNN-----STSNKKYSASALOTTASKVTFPAALQAEYTIETLRADNAGEVAA.GATNGR 554
Qy 473 KEKVTDSR-----KLVSXKANPDKLVQVQNKLPVTFVTDDQYDPPFGAATAIKEY 525
Db 555 EYKVIKDKAGNLAKNEIWNVAFNEDKORVISTVT-NAKFVDTD----- 597
Qy 526 LPTGVVABGG-----LDVVTTDSG-----IGTKT-----IGVTGNDVGGTV--- 564
Db 598 -PDTAVYFTGDKAKOISVKTDKGEATFVIGSDTVNDVATPIAMIDINTSDAKQGDLDG 656
Qy 565 -----HFG-----NGNG-----ATLGSLYVNVTEGNAFKNFELVSKVGQ----- 599
Db 657 EPKAVAP-SYFQAPYLDGSAIKAYKKSNDLNKAVTFKDGSETAVFAAELVNSGKKVTGTS 716
Qy 600 -----YGO-SPTKLDLV-----STTVEYQLSKVTSRVSVDPENLEGYEVESKNL 645
Db 717 IKATYTYTNTGANDIKVDNQVISPNSRYTVTYEATLSSTGTGVTIPAKNLEVTSDGKTT 776
Qy 646 AVADAKIVGNKVVVTG---KTPGKVDIHLTKNGATAG-KATVEIVQETIAIKSVNFKPVQ 701
Db 777 AV-----KVIATGIAVNTDGK-DYAFTAKEATATTAINEVPN-----SYTGVA 819
Qy 702 TE-NFVEKKINGTVLEKSNLDDIVKGINLTKETQHKVRVYKSGAQGL--YLRNG 758
Db 820 TQFNATDSGNSNSIWFAGKN-----PVKYAGVSGTKYKFGANG 859
Qy 759 DAVENAGDVKLGVDVTVSQTSDSALPNFKADLYDLTITTKYTDKG---TLAVEKVLKDKVIT 815
Db 860 NEVFGEA-----AWEALLTQATEGQKQVITISNV--DGDVTI 894
Qy 816 SEI 818
Db 895 FKV 897

RESULT 8
US-10-282-122A-46594
; Sequence 46594, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
```

```
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46594
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Bacillus anthracis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (161)..(161)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (165)..(165)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (180)..(180)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (184)..(184)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (185)..(185)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (303)..(303)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-46594
```

```
Query Match 8.8%; Score 369; DB 12; Length 408;
Best Local Similarity 31.0%; Pred. No. 6.5e-18;
Matches 148; Conservative 57; Mismatches 172; Indels 100; Gaps 21;

Qy 3 KSPDPVAGHWAEGSNLYLVKGAITGKPDGTGYPTESIDRASAAVIFTKILNLPVDENA 62
Db 22 KFPDVOQGGAEYSINYLAEKGAVTGNEKGMFEGKEITRAEAATMMAKILNLPIDKNA 81
Qy 63 QPSFKDKNINWSKYTAAVEKAGVVGDKGKENVYPEGKIDRASFAFMLVSNLKDKNYNG 122
Db 82 KPSYADSQKHATPIIAAVEKAGVVGKGTG-NGFEPDGKIDRVSMASLLVYAEYKLESKNR 140
Qy 123 -ELVTTTFEDLDH-WGEEKANIL-INLGISVGTGGKWEPNKS-VSRAEAAQFIALTDKKY 178
Db 141 PPOPINFKFKHKKRGEKRYFKXSWFFVGGGQRGXENFXKAEPEVHIKRPD--- 197
Qy 179 GKXDNAQAYVTDVKVSEPTKLTITGTGLDKLSADDVTLGDKKAVAIEASTDGTSAVVTLG 238
Db 198 -----SLKVGNNPLV-----EKVVIIDPGH-----G 217
Qy 239 GKVAPNKDLTVKYNQGSFVTKFVYVKKLAVKLTFTDDDRAGQAIAFLNDEKGN-ADVE 297
Db 218 GFDPGNPGQGV-----ESKIVFDT-SLRQKLL---EKNTPLKALLTREENGPGSNK 267
Qy 298 YLNLANHDYKFAVANNIDGSPANIFEGETATSTTGKLVGIKQGDYKVEYQVTKRGGLTVS 357
Db 268 NESLANR-VKFGQEN---NADIFVSIHANSEKHDGHGFKNVYK---KSKRGEETQI 318
Qy 358 NTGIIITVKNLDTTPASAIKN-VVPALDADNDGVVNYGSKLSGKDFALNSQNLVVGKASLN 416
```


Db 319 E-----KSEVLAKIQRNVEALHTRDRKI-----KD-----DHSLYVNNNTVP 359
Qy 417 KLVATAGEDKVPDPGSSISIKSSNHHIISVNNVITAAEAGEATLTI---KVGDTVK 470
Db 360 AVTELEAFIDINDIDNGKLATESGRQ-----IAAEAVYAGILDYEWKGFVSK 407
RESULT 9
US-09-117-447-6
; Sequence 6, Application US/09117447
; Patent No. US20020168728A1
; GENERAL INFORMATION:
; APPLICANT: LUBITZ, Werner
; APPLICANT: SLEYTR, Uwe
; APPLICANT: KUEN, Beatrix
; APPLICANT: TRUPPE, Michaela
; APPLICANT: HOWORKA, Stefan
; APPLICANT: RESCH, Stepanka
; APPLICANT: SCHROLL, Gerhard
; APPLICANT: SARA, Margit
; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF S-LAYER PROTEINS
; FILE REFERENCE: 100564-08013
; CURRENT APPLICATION NUMBER: US/09/117.447
; PRIOR FILING DATE: 1998-12-02
; PRIOR FILING DATE: PCT/EP97/00432
; PRIOR FILING DATE: 1997-01-31
; PRIOR FILING DATE: DE/196 03 649.6
; PRIOR FILING DATE: 1996-02-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 6
; LENGTH: 921
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-117-447-6

Query Match 8.7%; Score 367.5; DB 9; Length 921;
Best Local Similarity 24.9%; Pred. No. 2.9e-17;
Matches 220; Conservative 121; Mismatches 335; Indels 209; Gaps 47;
Qy 4 SPDPVAGHWAEGSYLVKGAITGPKDGTGPTSIDRASAAYFTKILNLPVDENAG 63
Db 33 SFTDVAPQY--KDAIDELVSTGATKGTETKFGYDEITRLDAVILARVLKLDVNDKD 90
Qy 64 PSFKDAKNIWSSKVIAAVEKAGVYVKGDKENFYEPGKIDRASPAFMLVSAYNLKOKVNGE 123
Db 91 AGFTDVPKD-RAKYNALVEAGVLNGKAPGKFGAYDPLTRVENAKIANKYKUK---ADD 146
Qy 124 LVTTFFDLDRHGE-----EKANILINLIGISVGTGKWEKPKSVSRAEAAQFIATLD 175
Db 147 VKLPFTDVNDTWAPYKALYKYEVTKRLKHQAQSVHT-----KNITLRDPAQFVY--- 196
Qy 176 KKYGKDNAGAYVTDVKSVEPTKLTLTGTGLDKLSADDVT-LEGDKAVAEIAST---DGT 231
Db 197 ----RAVNIINAVPIEVTAVNSTVTKVTNTQIADVDFNPAIDNGLTIVTKATLSRDKK 252
Qy 232 SAVVTGKGVAPNKDLTVK---VKN-----QGFVTKFVYEVKLAVEKLTEDDDRAGA 282
Db 253 SVEVVNKPFRNQEIYITATGINKLGETAKELTGKFWVSQDVAVTVALNSSLKVGEE 312
Qy 283 IAFKLNDEKNADVVEYLNLANHVKFVANNLDGSPANI---FEGGEATSTTGKLAUGIKOG 340
Db 313 SGLTVKQDQK-----DVVGAKVELTSNTNIVVVSSEGSVSAAKY-TAVKFG 360
Qy 341 DYKVEOVTKRGGLTVSNTGIIITVKNLDTPASAIKNVVFALDADNDGVNYSKLSGKDF 400
Db 361 TADVTAKVTLPDGVVLNTFEKVTVEPV---QVQNGFTL-VDN-----LSNA--- 405
Qy 401 ALNSQNLVWGEKASLNKLVIATIGEDKV-----DPGS-----TSIKSSNHHIIS 445
Db 406 ---PONTVAENKA--EKVTSNFAGETKVAVYDNTKNGDPTKVPDFKDATVRSNLPIAT 460

Qy 446 VVNN-----YITAEA--AGEATLTIKVGDTVK-----DVK-----FKVTTDSRLKLSVK 487
Db 461 RAINGSELLVITANQSGKASFEVTLKONTKRTFTVDVKKDPVLQDIKYDATSVKLSDEA 520
Qy 488 ANPDKLQVQNKILPVFTVTTDQYGDGPGANTAAIKVELPKTGVVAEGLDVVTTDSGSI 547
Db 521 VGGGEVGVNQKTIKVSAN--DQYG-----KEI--RFGT---KGRVTVTTTTEGL 563
Qy 548 GPKTIGVTGNDVGEQTVHFQNGNGATLGSLYVNVTEGNVAFKPFELSVKVGQYGSPDTK 607
Db 564 VIKNVN-----SDNTIDFDSGNSAT-----DQFVVA-----TK 592
Qy 608 LDLNVSTTVEYQLSKYTSRDRVYSDPENLEGEVSESKNLAVADAKIVGNKVVVGTGTPGV 667
Db 593 -DKIVNGKEVYFKNASD---TPTSTKTTITVNVVNVK-ADATPVGLDIV---APSKI 643
Qy 668 DIHLTKNGATAGKATVEIVQ-ETIAIKSVN-----FKPVQTE---NFEKKINIGTV 715
Db 644 DVNAPN---TASTADVDFINFESVEIYTLDSNGRRQKVVTTATLTGTTKXKKVN-GNV 699
Qy 716 LELEKSNLDDIVKGINLTKETQHKVVRVYKSGAE-----QGLYLDNRNGDAVFNAGDYKL 769
Db 700 LQFKGNE-----ELTSTSSSTGNDVGTAGMTKRIIPGK-YI--NSASVPASATVAT 748
Qy 770 GDTVYS-QTSDSALPNFKADLYDTLTTKYTDKGTIVFKVLKDKV 813
Db 749 SPVTVKLNSDNDL-----TFELIFGVIDPTQLV-----KDEDI 783
RESULT 10
US-10-092-880-9
; Sequence 9, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092.880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1599
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-9
Query Match 6.1%; Score 257.5; DB 13; Length 1599;
Best Local Similarity 22.0%; Pred. No. 4.6e-09;
Matches 183; Conservative 104; Mismatches 328; Indels 215; Gaps 38;
Qy 16 GSI-----NYLVDKGA-----ITGKPDGTGPTSIDRASAAYFTKILN---L 56
Db 891 GSIINIAGNLTVSKGANLQAITNTFNVAGSFNDNGASNISIAAGGAK--FKDINNTSSL 948
Qy 57 PVDENAOFPKDAKNIWSSKVIAAVEKAGVYVKGDKENFYEPGKIDRASPAFMLVSAYNL 116
Db 949 NITNSDITVR-----TIKGNISNKGDLNI-----IDKSDAEIQIGG-NI 990
Qy 117 KDKNGELVTTFFDLHGWBEKANILINLIGISVGTGKWEKPKSVSRAEAAQFIATLDK 176
Db 991 SKQ-EGNLTI-----SDKVNITNQITIKAGVEG----- 1018
Qy 177 KYGKDNAGAYVTDVKSVEPTKLTLTGTGLDKLSADDVTLEGDKAVAEIASTGTSVAT 236
Db 1019 --GRSDSSEA-----ENANLTIQTKEL-KL-AGDLNISGFNKAEITAKNGSDLTIGN 1066

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:30:55 ; Search time 23 Seconds
(without alignments)
1869.758 Million cell updates/sec

Title: US-09-844-281-1
Perfect score: 4202
Sequence: 1 AGKFPDPVAGHWAAGSINY.....ITSEIGQAVHVNLPNL 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/2/iaa/5A_COMB.pdp.*
2: /cgm2_6/ptodata/2/iaa/5B_COMB.pdp.*
3: /cgm2_6/ptodata/2/iaa/6A_COMB.pdp.*
4: /cgm2_6/ptodata/2/iaa/6B_COMB.pdp.*
5: /cgm2_6/ptodata/2/iaa/PTCUS_COMB.pdp.*
6: /cgm2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	374	8.9	1222	2	US-08-682-517-15
2	374	8.9	1252	2	US-08-682-517-9
3	367	8.7	920	4	US-09-463-402-6
4	362.5	8.6	921	4	US-09-889-572-4
5	257.5	6.1	1338	2	US-08-728-470-9
6	257.5	6.1	1338	3	US-08-719-641-9
7	257.5	6.1	1599	2	US-08-617-697-9
8	251.5	6.0	1529	2	US-08-728-470-10
9	251.5	6.0	1529	3	US-08-719-641-10
10	249.5	5.9	1600	2	US-08-617-697-10
11	249.5	5.9	2353	3	US-09-377-155-33
12	249.5	5.9	2353	3	US-08-913-942-4
13	249.5	5.9	2353	4	US-09-669-974-33
14	249.5	5.9	2353	4	US-09-797-862-33
15	247.5	5.9	2354	4	US-09-268-347-47
16	246	5.9	1180	4	US-09-206-942-65
17	246	5.9	1188	4	US-09-206-942-63
18	244.5	5.8	1912	1	US-08-409-995-4
19	244.5	5.8	1912	3	US-08-685-467-4
20	242.5	5.8	1220	4	US-09-206-942-28
21	242.5	5.8	1226	4	US-09-206-942-26
22	240.5	5.7	2411	4	US-09-268-347-36
23	237.5	5.7	2314	4	US-09-268-347-49
24	235.5	5.6	2048	4	US-09-268-347-48
25	235	5.6	1004	4	US-09-206-942-57
26	235	5.6	1010	4	US-09-206-942-55
27	229	5.4	969	4	US-09-206-942-32

28	229	5.4	975	4	US-09-206-942-30	Sequence 30, Appl
29	227.5	5.4	1073	4	US-09-206-942-49	Sequence 49, Appl
30	227.5	5.4	1079	4	US-09-206-942-47	Sequence 47, Appl
31	227	5.4	2504	4	US-09-328-352-5821	Sequence 5821, Ap
32	222	5.3	1833	4	US-08-621-944A-4	Sequence 4, Appli
33	222	5.3	1833	4	US-08-945-567D-4	Sequence 4, Appli
34	222	5.3	1992	4	US-08-621-944A-3	Sequence 3, Appli
35	222	5.3	1992	4	US-08-945-567D-3	Sequence 3, Appli
36	219.5	5.2	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
37	217.5	5.2	1005	4	US-09-206-942-41	Sequence 41, Appl
38	217.5	5.2	1011	4	US-09-206-942-39	Sequence 39, Appl
39	216.5	5.2	1095	4	US-09-206-942-69	Sequence 69, Appl
40	216.5	5.2	1536	1	US-08-038-682-2	Sequence 2, Appli
41	216.5	5.2	1536	1	US-08-302-832-2	Sequence 2, Appli
42	216.5	5.2	1536	2	US-08-530-198-2	Sequence 2, Appli
43	216.5	5.2	1536	2	US-08-469-880-2	Sequence 2, Appli
44	216.5	5.2	1536	2	US-08-728-470-2	Sequence 2, Appli
45	216.5	5.2	1536	2	US-08-617-697-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-682-517-15
; Sequence 15, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-682-517-15

Query Match 8.9%; Score 374; DB 2; Length 1222;
Best Local Similarity 23.3%; Pred. No. 2.7e-19;
Matches 224; Conservative 130; Mismatches 345; Indels 264; Gaps 47;

QY	13	WAEGSTNYLVDKGAITGPKDGYGPTESIDRASAIVFTKILNLPVDENAOQSFKDAK-N 71
DB	12	YAKEAVQALVDQGVLCQGTNGNFNPLNVTAAQAEITKALEL--EANGDVNFKDVKAG 69
QY	72	INSSKYIAAVERKAGVYKGDKENFYPEGKIDRASPASMLVSAYNLKOKVNGELVTTFEDL 131
DB	70	AWYNSIAAVVANGIFEGVSATEFAPNKSILTRSEAAKILVEAFGLEGEAD---LSEFADA 126
QY	132	--LDHKGEBKANILNLGISVCT--GKWEPNKSVSRAE--AAOFIALTDKYGCKDNAQAY 187
DB	127	SOVKPWAKYLEIAVANGIFEGTDANKLNPNNSITRQDFALVFKETVTKVEGETEEAFA 186
QY	188	V-----TDYKVS---EPTKLTITGTGLDKLSADDDVTL--EGDKAVAI---EASTDGTSAVV 235
DB	187	VKAINTTVEVTFEEBVTNVQALNFKEGLEIKNASVQTNKKVVVLTTEAQTADKYEVL 246
QY	236	TLGKK-----VAPNKDITVKVKQCSFVKFVYEVKKLAVEKLIFFDDDDRAQOIAF 285
DB	247	TLDGETIGGFKGVAADVPTK---VELVSAVQGLQGVKQVAKVTVAEGQSKAGIPVTF 303
QY	286	-----KLNDKGNADVBYLNLNHDVKFVA-----NNLDGSPANIFEKG 324

Db 304 TVPGNNNDGVVPTLTGALTNEEGTATVSYTRYKEGTDDEVTAATGDRSKFSLGYVFWGV 363
QY 325 EATSTTGKLVG--IKQG--DYKVEVQVTKRGGLTVS---NTGIIIVKVL--DTPASAI 374
Db 364 DTILSVEVTTGASVNNGANKTYKYTKNPKTGKPEANKTFNVGFVENMMVTSKVNANAT 423
QY 375 KNVFALDADNDGVVNYGSKLGDFAINSONLVVGEKASLNKLVIATIGEDKVPDPGSI 434
Db 424 VNGVKALQLSN-----GTALDAAQITTDGK-----GEATFTVSGTNAAVTPVYV 467
QY 435 SIKSNHGIISVNNYITAB-----AAGEATLTIKV-----GDV-----TKDV 472
Db 468 DLHSTNN---STSNKKYSASALQTTASKVTFPAALQAEYTIELTRADNAGEVAAGTNGR 524
QY 473 KFKVTTDSR-----KLVSUKANPDQLQVQNKTLPTVFTVTTDQYDPPFGANTAAIKEY 525
Db 525 EYKVIKDKAGNLAKNEIWNVAFNEDKDVISTVT--NAKFVDTD----- 567
QY 526 LPKTGVVAEGG---LDVVTTDSGS-----IGTKT-----IGVTGNDVGEQTV--- 564
Db 568 -PDTAVYFTGDKAKQISVKTNDKGEATFVIGSDTVNDYATPIAWIDINTSDAKQGDLDG 626
QY 565 -----HFQ-----NGNG-----ATLGSLYVNVTEGNAFKNFELVSKVGQ----- 599
Db 627 EPKAVAPISYFQAPVLDGSAIKAYKKSDLNKAATKFDGSETAVFAELVNQSGKKVTGTS 686
QY 600 -----YGO--SPDTKLDLV-----STTVEYQLSKYTSRDVSDPENLEGYEVESKNL 645
Db 687 IKKATYTIINTGANDIKVDNQVISPNSRYTVYEATLSSTGTVITPAKNLEVTSDGKTT 746
QY 646 AVADAKIVGNKVVTVG---KTPGKVDIHLTKNGATAG--KATVEIVQSTIAIKSVNPKPVQ 701
Db 747 AV-----KVIATGIAVNTDGG--DYAFTAKEATATATATNEVPN-----SYTGVA 789
QY 702 TE-NFVEKKINGITVLELEKSNLDDIVKGINLTKEQHKVRVVKSGAEQKGL--YLDNRG 758
Db 790 TQFNTADSGSNSNIWFAGKN-----PVKYAGVSGKTYKYFGANG 829
QY 759 DAVFNAGDVKLGDVTVTSQSDSALPNFKADLYDILTITKYTDGK---TLVFKVLKDKDVIT 815
Db 830 NEVFGEA-----AWEALLTQYATEGOKVTIISYNV--DGDVTV 864
QY 816 SEI 818
Db 865 PKV 867

RESULT 2
US-08-682-517-9
; Sequence 9, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/682,517
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-682-517-9

Query Match 8.9%; Score 374; DB 2; Length 1252;
Best Local Similarity 23.3%; Pred. No. 2.8e-19;
Matches 224; Conservative 130; Mismatches 345; Indels 264; Gaps 47;
QY 13 WAEGSINYVDKGAITGKPDGTGYGPTESIDRASAAVITKILINLPVDENAPSPFDKAK-N 71
Db 42 YAKEAVQALVDQGVTOGDTNGNFNPLNTVTRAQAABIFTKALEL--EANGDVNFKDVFKAG 99
QY 72 IWSSKYIAAVEKAGVVGKDGKGFPEGKI DRASPASMLVSAYNLKDKVNGELVTTFFDL 131
Db 100 ANYNSIAAVVANGIFEGVSATEFAPNKSILTRSEAAKILVEAFGLEGEAD--LSEFADA 156
QY 132 --LDHWGEKANILINLIGSVGT--GKWEPEKNSVSRAB--AAQFIALTITKYYKKKDNAAQY 187
Db 157 SQVKPMWAKYIEIAVANGIFEGTDANKLNPNNSITRQDFALVFKRTVDKVEGETPEEAAAF 216
QY 188 V-----TDVKVS---EPTKLTGTGLDKLSADDDVTL--EGDKAVAI---EASTDGTSAVV 235
Db 217 VKAINNTTVEVTFEESEVINVQALNFKIEGLEIKNASVKQTKNKVVVLTTEAQTADKEVYL 276
QY 236 TLGGK-----VAPNKDLTVKVKQNSFVTKFVYEVKLVAVKELTFDDDDRQAQATF 285
Db 277 TLDGETIGGFGVAAAVVETK---VELVSSAVQGKLGQEVKQVQKVTVAEGOSKAGIPVTF 333
QY 286 -----KLNDKGNADVEYLNLANHDKVFA-----NNLDGSPANIFEGG 324
Db 334 TVPGNNNDGVVPTLTGALTNEEGTATVSYTRYKEGTDDEVTAATGDRSKFSLGYVFWGV 393
QY 325 EATSTTGKLVG--IKQG--DYKVEVQVTKRGGLTVS---NTGIIIVKVL--DTPASAI 374
Db 394 DTILSVEVTTGASVNNGANKTYKYTKNPKTGKPEANKTFNVGFVENMMVTSKVNANAT 453
QY 375 KNVFALDADNDGVVNYGSKLGDFAINSONLVVGEKASLNKLVIATIGEDKVPDPGSI 434
Db 454 VNGVKALQLSN-----GTALDAAQITTDGK-----GEATFTVSGTNAAVTPVYV 497
QY 435 SIKSNHGIISVNNYITAB-----AAGEATLTIKV-----GDV-----TKDV 472
Db 498 DLHSTNN---STSNKKYSASALQTTASKVTFPAALQAEYTIELTRADNAGEVAAGTNGR 554
QY 473 KFKVTTDSR-----KLVSUKANPDQLQVQNKTLPTVFTVTTDQYDPPFGANTAAIKEY 525
Db 555 EYKVIKDKAGNLAKNEIWNVAFNEDKDVISTVT--NAKFVDTD----- 597
QY 526 LPKTGVVAEGG---LDVVTTDSGS-----IGTKT-----IGVTGNDVGEQTV--- 564
Db 598 -PDTAVYFTGDKAKQISVKTNDKGEATFVIGSDTVNDYATPIAWIDINTSDAKQGDLDG 656
QY 565 -----HFQ-----NGNG-----ATLGSLYVNVTEGNAFKNFELVSKVGQ----- 599
Db 657 EPKAVAPISYFQAPVLDGSAIKAYKKSDLNKAATKFDGSETAVFAELVNQSGKKVTGTS 716
QY 600 -----YGO--SPDTKLDLV-----STTVEYQLSKYTSRDVSDPENLEGYEVESKNL 645
Db 717 IKKATYTIINTGANDIKVDNQVISPNSRYTVYEATLSSTGTVITPAKNLEVTSDGKTT 776
QY 646 AVADAKIVGNKVVTVG---KTPGKVDIHLTKNGATAG--KATVEIVQETIAIKSVNPKPVQ 701
Db 777 AV-----KVIATGIAVNTDGG--DYAFTAKEATATATATNEVPN-----SYTGVA 819
QY 702 TE-NFVEKKINGITVLELEKSNLDDIVKGINLTKEQHKVRVVKSGAEQKGL--YLDNRG 758
Db 820 TQFNTADSGSNSNIWFAGKN-----PVKYAGVSGKTYKYFGANG 859
QY 759 DAVFNAGDVKLGDVTVTSQSDSALPNFKADLYDILTITKYTDGK---TLVFKVLKDKDVIT 815
Db 860 NEVFGEA-----AWEALLTQYATEGOKVTIISYNV--DGDVTV 894
QY 816 SEI 818
Db 895 PKV 897

341 DYKVEQVTKRGELTVSNTGLITVKNLDTASAIAKVVVFALDADNDGVVNYGSKLGGKDF 400
361 TADVTAKVTLDPGVVLTITFRVTVTEVPV---QVQNGFTL-VDN-----LSNA-- 405
401 ALNSQNLVVGKASLKLAVIAGBDKV---DPGS-----ISIKSNHGIIS 445
406 ---PQNTVAFNKA---EKVTSFAGETKTVAMVDTKNGDPETKVPDFKDATVRSINPIAT 460
446 VVNN---YITAEA---AGEATLTIKVGDTK---DVK-----FKVTTDSRLYSVK 487
461 AAINGSELLVTANAGOSKASPEVTKNTKETFTVDVVKDPVLODIKVDATSVKLSDEA 520
488 ANPKDLQVQNKTLPTFTVTTDQYQDPFGPAGTAALKEVLPTKVVVAEGGLDVVTTDSGI 547
521 VGGGEVGVNQTKIVSAV---DQYG---KEI---KFGT---KGVTVVITNTEGL 563
548 GTTKTIGVTGNDVGEVTVHFQNGNGATLSLVV---NVTEGNVAFKPELVV----- 595
564 VIKNVN-----SDNTIDFSGNSATDQFVVVATKDKIUNGKVEVYKFNASDTPTPTSTK 617
596 -----KVGQYQSQPDTKLDLNVSTTVVEYQLSKYTSDRVYSDPENLEGYEVE 641
618 TITVNVVNVKADATPGLVDIVAP---SEIDVNAENTAS-----TADVDFINFESVEIYTL 670
642 S-----KNLAVADAKIVGNK---VVVTGK-----TPGKVDIHLTKNGAT--- 677
671 SGNRLKKVTPATTTLVGTNDYVEVNGVLOPKGNDELTLTSSSTVNVVDVADGITKRI 730
678 ---AGKATVEIVQETIAIKVNFKPVOTEN---FVEKKINIGTVLELEKSNLDDIVKGI 730
731 PVKYNASVP-ASATVATSPVTVKLNSSDNDLTFFELIFGVDPQLVKD---EDINEFI 787
731 NITKETOH-----KVRVVKSGAE-----QKLYLDNR-----G 758
788 AVSKAAKNDGVLNKLPLVTVKDSAGKVIPTGANVYGLNHDATNGNIWFDEQAGLAKFS 847
759 DAVEN-----AGDKVLDGVTSQTSALPNFKADLYDTLTTPKYTDKG-----TLVFK-- 806
848 DVHFDVDFSLANVVKVGTGTVSS-----PSLSDAI---QLTNSGDVAFSLIVIKSI 896
807 VLKDKDVITSEIGSQVHVNV 827
897 YVKGADKDDNNLLAAPVSVNV 917

RESULT 5
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 6.1%; Score 257.5; DB 2; Length 1338;
Best Local Similarity 22.0%; Pred. No. 1.8e-10;
Matches 183; Conservative 104; Mismatches 328; Indels 215; Gaps 38;
QY 16 GSI-----NYLVDKGA-----ITGKPDGTGPTESIDRASAIVFTKILN---L 56
DB 631 GSIINAGNLTVSKANLQAITNYTFNVAGSFDDNNGASNISARGGAK--FKDINNTSL 688
QY 57 PVDENQOPFKDAKNWSSKYIAAVEKAGVKGDKGFENYVEPKGIDRASFASMLVSAYNL 116
DB 689 NITNEDTIVR-----TIKGNISNKSGLNI-----IDKSDAEIQIGG-NI 730
QY 117 KDKVNGELVTFPEDLDHMGEEKANLILNIGISVGTGKWEKPNKSVSRABAAQFIALTDK 176
DB 731 SQK-EGNLNTIS-----SDKVNITNQITIKAGVEG-----758
QY 177 KYGKKNQAQVVDVVKVSBPTKLTGTGLDKLSADDVLTGDKAVAIASDTGTSAVVT 236
DB 759 --GRSDSSEA-----ENANLTIQTKEL-KL-AGDLNISGNKAEITAKNGSDLTIGN 806
QY 237 LGGKVAPNKLTV-KVKNSQSFVT-----KFVVEVKLAVEKLTFFDDDRAGQAIAFKLNDE 290
DB 807 ASGNADAKKVPDKVKDKSKISTDGHNVTLNSEVKTSGNSNAGNDNSTGLTISAK--- 862
QY 291 KGNADVEYLNLAHDVKFVANNLDGSPANIF--EGGEATSTTCKLAVGKQGDYKVEVQV 348
DB 863 ----DV---TVNNNVTSKHTINISAAAGNVTTKEGTTINATTGSEVTAQNG----- 907
QY 349 TKRGGLTVSNTGITVKNLDTPASAIKNVVVFALDADNDGVVNYGSKLGGKDFALNSQNLV 408
DB 908 TIKGNITSONVTVTATENLVTTENAVINA-----TSGTVNISTKTGDIKGGIESTSGN 960
QY 409 VGEKASLNKL-VATIAGEDKVPDPGSIKSNHGIISVVNNVITAEAGEALTIKVG 467
DB 961 VNITASGNTLKVSNITQDVTV-----TADAGALTTTAGTSTIATTGNANITTKTG 1012
QY 468 VTKDVK-----FKVTTDSRLYSVKANPKDLQVQNKTLPTV--FVTTDQY 511
DB 1013 INKVESSSGSVTLVATGATLAVNIGNTVTTTADSGKLTSTVSGTINGNSVTTSSQS 1072
QY 512 GDFPGANTAAKEVLPKPTG-----VVAEGGLDVTTDSGISGTGKIGTVGNDVGE 562
DB 1073 GDIEGTISGNTVNVVNTASTGDLTIGNSAKVEAKNGAATLTAESGKLTQ---TGS 1125
QY 563 TVHFQNGGATLGLSVNVVTEGNVAFKPELVSKVQYQSGSDTKLDLNVST-TVEYQLS 621
DB 1126 ----TSSNGOTTLTAKDSSAGNINAVNVL-NTTGLTTGDSKINATSGILTIKAKDA 1180
QY 622 KY-----TSDRVYSDPENLEGYEVEKSNLAVADAKIVGNKVVVTKTPEKVDIH---LTKN 674

Db 1181 KLDGAASGDRVTVNATVNSG-----SGN-----VTAKTSSSVNITGDLNTIN 1222

Qy 675 G-----ATAGKATVEIVQETIAIKSVNFKPVQTFENFVEKKINIGTVLEKSNLDDIVK-G 729

Db 1223 GLNIISSENGRNTVLRGKEIDVKVIQPGVASVEVIEAKRVLEKVKDLSDEBERETLAKLG 1282

Qy 730 IN-----LTKETQHKVRVVKSG---AEQKGLYDRNGDAVFNAGD 766

Db 1283 VSAVRFEVPNNAITVNTQNEFTTKPSSQVTISEGK-----ACFSSGN 1324

RESULT 6

US-08-719-641-9

Sequence 9, Application US/08719641

Patent No. 6218141

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/719,641

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 1038-625

REFERENCE/DOCKET NUMBER: 22,651

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1338 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-719-641-9

Query Match 6.1%; Score 257.5; DB 3; Length 1338;

Best Local Similarity 22.0%; Pred. No. 1.8e-10;

Matches 183; Conservative 104; Mismatches 328; Indels 215; Gaps 38;

Qy 16 GSI-----NYLVKGA-----ITGKPDQTYGPTESIDPRASAAVFTKILN---L 56

Db 631 GSIINIAGNLTVSKGANLQAITNTVFNAGSFDNNGASNTSIARGGAK--FKDINNTSSL 688

Qy 57 PVDENAOFSFKDAXNWSKYLAAVERKAGVVGKGFYFEGKIDPRASFASMLVSNYL 116

Db 689 NITNSDITYR-----TIKGNISNKGDLNI-----IDKSDAEIQIGG-NI 730

Qy 117 KDKVNGELVTFPEDLLDHNGBEKANILINLQISVGTGKGWEPKNKSVSRABAAQFIALTDK 176

Db 731 SOK-EGNLTIS-----SDKVNITNQITIKAGVEG----- 758

Qy 177 KYGKDNAAQAVVDVKVSEPTKLTGTGLDKLSADDVTLBGDKAVAEASTDGTSAVVT 236

Db 759 --GRSDSEA-----ENANLFIQTKEL-KL-AGDLNISGFNKAETAKNGSDLTIGN 806

Qy 237 LGKQVAPNKDLTV-KVKNQSFVT-----KFVVEVKLAVEKLTFPDDDRAGQAIAPKLANDE 290

Db 807 ASGGNADAKVTFDKVDSKISTDGHNVTLNSEVKTSGSSNAGNDNNTGLTISAK----- 862

Qy 291 KGNADVEYLNLANHDVKFVANNLDGSPANIF--EGGEATSTTGKLAVGIKQGDYKVEQV 348

Db 863 ----DV---TVNNVTSHKTINISAAAGNVTKETGINATIGSEVETQAG----- 907

Qy 349 TKRGGLTVSNTGIIITVKNLDPASAKNVFALDADNDGVVNYGSKLGGKDFALNSONLV 408

Db 908 TIKGNITQNVTVTATENLVTTENAVINA-----TSGTVNISTKTGDIKGGIESTSGN 960

Qy 409 VCEKASLNKL-VATTIAGEDKVDVDPGSIKSNHGIISVVANNYITAEAGATLTIKVGD 467

Db 961 VNITASGNTLKVSNTGQDVT-----TADAGALTITAGSTISATGNANITTKGD 1012

Qy 468 VTKDVK-----FKVTTDSRLKSVKANPKQLQVQNKTLPLVT--FVTTDQY 511

Db 1013 INKVESSSSGSVTLVATGATLAVGNISGNTVITADSGKLTSTVGSTINGNSVTTSSQS 1072

Qy 512 GDFPGANTAAIKVLPKTG-----VVAEGGLDVTTDSGSGITGKTIGTGNDDVSEG 562

Db 1073 GDIEGTISGNTVNTASTGDLTIGNSAKVEAKNGAATLTABSGKLTTQ-----TGSSI--- 1125

Qy 563 TVHFQNGNGATLGLSVNVNTEGNVAFKFNELVSKVGQYQSGPDTKLDLNVST-TVEYQLS 621

Db 1126 ----TSSNGQTLTAKDSSIAGNINAAVTL-NTGTTLTTGDSKINATSGTLTINAKDA 1180

Qy 622 KY-----TSDRVSDPENLEGYEVESKNLAVADAKIVGNKVVTGKTPGKVDIH---LTKN 674

Db 1181 KLDGAASGDRVTVNATVNSG-----SGN-----VTAKTSSSVNITGDLNTIN 1222

Qy 675 G---ATAGKATVEIVQETIAIKSVNFKPVQTFENFVEKKINIGTVLEKSNLDDIVK-G 729

Db 1223 GLNIISSENGRNTVLRGKEIDVKVIQPGVASVEVIEAKRVLEKVKDLSDEBERETLAKLG 1282

Qy 730 IN-----LTKETQHKVRVVKSG---AEQKGLYDRNGDAVFNAGD 766

Db 1283 VSAVRFEVPNNAITVNTQNEFTTKPSSQVTISEGK-----ACFSSGN 1324

RESULT 7

US-08-617-697-9

Sequence 9, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617,697

```

; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-9

Query Match      6.1%; Score 257.5; DB 2; Length 1599;
Best Local Similarity 22.0%; Pred. No. 2.4e-10;
Matches 183; Conservative 104; Mismatches 328; Indels 215; Gaps 38;

QY      16 GSI-----NYLVDKGA-----ITKPDGTGYGPTESIDRASAAVFTKLNL--L 56
Db      891 GSIINTAGNLTYSKGANLOAITNTFNAGSPDNNGASNIARGGAK--FKDINNVTSSL 948

QY      57 PVDENAQPSFKDAKNISWSSKYIAAEKAGWKGDGFENFYPEGKIDRFASFASMLVSAYNL 116
Db      949 NITTNSDTTYR-----TIKGNISNKSGDLNI-----LDKSDAIEIGG-NI 990

QY      117 KKKVNGELVTTREDDLLDHWEKEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTOK 176
Db      991 SQK-EGNLAIS-----SDKVNITNQITIKAGVEG-----1018

QY      177 KYGKDNDAAVTDVKVSPBTKLTGTGLDKLSADDVTLGGDKAVATEASTDGTCSAVVT 236
Db      1019 --GRSDSSA-----ENANLTIQTVEL-KL-AGDLNISGFPAEITAKNGSDELIGN 1066

QY      237 LGSKVAPNKDLTV-KYKNOSFVT-----KFVYEYVKKLAVEKLTPDDRAGQAIAFKLNDE 290
Db      1067 ASGCNDAKAVTFDKVKDSKISTDGHNVTLNSEVKTSGSSNAGNDNSTGLTISAK--- 1122

QY      291 KGNADVEYLNLANDHVKFVANILDSPANIP--EGGEATSTTGKLVAGIKQGDYKVEVOV 348
Db      1123 ----DV---TVANNVTSHKTNISAAAGNVTTKEGTINATTGSEVETVAQNG----- 1167

QY      349 TKREGGLTVSNTGIITVKNLDPASPAAIKXNVFALDADNDGVNYSKLSGKDFALNSQLV 408
Db      1168 TIKNITSQNVTATENLVTTENAVINA-----TSGTVINSTKTDIKGGIESISGN 1220

QY      409 VGEKASLNKL-VATIAGEKVDVPGSISIKSNHGIIISVVNNYTAEAGBATITIKVGD 467
Db      1221 VNITASGNLTKVSNITGQDVT-----TADAGALLTTAGSTISATGNNANITTKTGD 1272

QY      468 VTCDVK-----PKVTTDSRLKSVSKANPDKLQVQNKPLPV--FVTTDOY 511
Db      1273 INKVBESSGSVTLVATGATLAVGNISGNTVTTITADSGKLSTVGSTINGTNSVTTSSQS 1332

QY      512 GDPFGANTAARKVLPEKTC-----VVAEGGLDVVTPDSSIGTKTIGTIVGNDVGE 562
Db      1333 GDIEGTISGNTVNVTASTDGLTAGNSAKVEAKNGAATLTAE SGKLTQ----TGSSI--- 1385

QY      563 TVHFQNGCATLGLSVNVTGNAVAFKPFELVSKVGYQGSPDYTKLDLNVST-TVEYQLS 621
Db      1386 -----TSNCGQTLITAKDSSIAGNINAAVNTL--NTTGTTLTTTGDSKINATSGLTLINAKDA 1440

QY      622 KY-----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVVTGKTPGVKDIH---LTKN 674

```

115 NLKDKVNGELVTFEDLDHMGEEKANILNLGISVGTGKWEPEKNSVSRAAQFIALT 174
888 TOKTIINGNITNEKGLD--NIKNIKADAEIQIG--GNISQKEGNTLTSKVNITNQIT 942
175 DK--KYGKKNAQAVTVDKVSSEPTKLTCTGLDLSADVDVLEGDKAVAEASTDGT 231
943 IKAGVEGGRSDSEA-----EVANITIQTKEL--KL-AGDLNISGFNKAEITAKNGSD 992
232 SAVVTLGKGAAPNKDLTV-KVKNQSPVT-----KFVVEYVKLAVERKLTFFDDRAGQAI 285
993 LTIGNASGNADAKKVTDFKVDKSISTDGHVNLNSVKTSGNSAGNDNSTGLTISA 1052
286 KLANDEKGNADVEYLNLANHDKVFNALDGPANIF--EGGEATSTTKLAVGKQGDYK 343
1053 K-----DV---TVNNVTSHTKINISAAAGNVTTKEGTTINATTSVEVTAQNG--- 1098
344 VEQVTKRGGLTVSNTGIIITVKNLDTPASAIKXNVFALDADNDGVVNYGSKLSGKDFALN 403
1099 -----TIKGNITVTSAGNTLKVSNTITQGVTV-----TADAGALTTTAGTISATTGNANIT 1198
403 SONLVVGEKASLNKL-VATIAGEDKVDPGSGISIKSNHGIISVNNVITABAAAGEATLT 462
1147 STSGNVNITASGNTLKVSNTITQGVTV-----TADAGALTTTAGTISATTGNANIT 1198
463 IKVGDVTKDVK-----FKVTTDSRKLVSVKANPDKLVQVQNKLPVT--FV 506
1199 TKTGIDNGKVESSSGSVTLVATGATLVAGNISGNTVTTTADSGKLTSTVGSTINGTNSVT 1258
507 TTDQYDGPFGANTAAIKVLPKIG-----VVAEGGLDVVTTDSGSGIKTKTIGVTGN 557
1259 TSSQSGDIEGTISGNTVNTASTGDLTIGNSAKVEAKNGAATLTAESGKLTQ---TGS 1314
558 DVGEGTVHFQNGNGATLGLSVYVNTVEGNVAFKXNFELSVKVGQYQSGPTKLDLVST-TV 616
1315 SI-----TSSNGQTTLTAKDSSIAGNINAANVTL-NTTGTLTGTTGDSKINATSGTLTI 1366
617 EYQLSKY-----TSDRVYSDPENLEGEYVESKNLAVADAKIVGNKVVTGKTPGKVDIH-- 670
1367 NAKDAKLGAASGDRTVNATNASG-----SGN---VTKATSSSVNITGD 1408
671 -ITKNG-----ATAGKATVEIIOETTAIKSVAFKPVQTFENFVEKKINIGTVLEKSNLDD 725
1409 LNTINGLNIISNGRNTVRLRKEIDVKYIQGVASVEEVEAKRVLEKVKDLSDERET 1468
726 IVK-GIN-----LTKEQHVVRVVKSG---AEQGLYLDKNGDAVFNAGD 766
1469 LAKLGVSARFVEPNNAITVNTQNEFTTKPSSQVITSEK-----ACFSSGN 1515

RESULT 9

US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641

FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
NAME: Bekstreser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 6.0%; Score 251.5; DB 3; Length 1529;
Best Local Similarity 22.5%; Pred. No. 6.3e-10;
Matches 188; Conservative 121; Mismatches 345; Indels 181; Gaps 38;
QY 16 GSINLVLDKGAITKPDGTYGPTPE-----SIDRASAIVFIKIINLPV 58
DB 778 GNIN-ITNKANVTLQADTSNNTGLKRTLTILGNISVEGNLSLTGANIVGN--LSIAE 834
QY 59 DENAQSPFDKAKNI---WSKVIATAVE-KAGVVKGDGKENFYEPGKIDRASASMLVSY 114
DB 835 DSTFKGEASDNLNITFTTNGTANINIKQVVKLQGDIN--NKSGLNTTTNAS-----G 887
QY 115 NLKDKVNGELVTFEDLDHMGEEKANILNLGISVGTGKWEPEKNSVSRAAQFIALT 174
DB 888 TOKTIINGNITNEKGLD--NIKNIKADAEIQIG--GNISQKEGNTLTSKVNITNQIT 942
QY 175 DK--KYGKKNAQAVTVDKVSSEPTKLTCTGLDLSADVDVLEGDKAVAEASTDGT 231
DB 943 IKAGVEGGRSDSEA-----EVANITIQTKEL--KL-AGDLNISGFNKAEITAKNGSD 992
QY 232 SAVVTLGKGAAPNKDLTV-KVKNQSPVT-----KFVVEYVKLAVERKLTFFDDRAGQAI 285
DB 993 LTIGNASGNADAKKVTDFKVDKSISTDGHVNLNSVKTSGNSAGNDNSTGLTISA 1052
QY 286 KLANDEKGNADVEYLNLANHDKVFNALDGPANIF--EGGEATSTTKLAVGKQGDYK 343
DB 1053 K-----DV---TVNNVTSHTKINISAAAGNVTTKEGTTINATTSVEVTAQNG--- 1098
QY 344 VEQVTKRGGLTVSNTGIIITVKNLDTPASAIKXNVFALDADNDGVVNYGSKLSGKDFALN 403
DB 1099 -----TIKGNITVTSAGNTLKVSNTITQGVTV-----TADAGALTTTAGTISATTGNANIT 1198
QY 404 SONLVVGEKASLNKL-VATIAGEDKVDPGSGISIKSNHGIISVNNVITABAAAGEATLT 462
DB 1147 STSGNVNITASGNTLKVSNTITQGVTV-----TADAGALTTTAGTISATTGNANIT 1198
QY 463 IKVGDVTKDVK-----FKVTTDSRKLVSVKANPDKLVQVQNKLPVT--FV 506
DB 1199 TKTGIDNGKVESSSGSVTLVATGATLVAGNISGNTVTTTADSGKLTSTVGSTINGTNSVT 1258
QY 507 TTDQYDGPFGANTAAIKVLPKIG-----VVAEGGLDVVTTDSGSGIKTKTIGVTGN 557
DB 1259 TSSQSGDIEGTISGNTVNTASTGDLTIGNSAKVEAKNGAATLTAESGKLTQ---TGS 1314
QY 558 DVGEGTVHFQNGNGATLGLSVYVNTVEGNVAFKXNFELSVKVGQYQSGPTKLDLVST-TV 616
DB 1315 SI-----TSSNGQTTLTAKDSSIAGNINAANVTL-NTTGTLTGTTGDSKINATSGTLTI 1366

QY 617 EYOLSKY-----TSRVYSDPENLEGVEYESKNLAVADAKIVGNKVVTGKTPGVKVDH-- 670
DB 1367 NAKADLGAASGDETVVNAASG-----SGN-----VIAKTSVVNTG 1408
QY 671 -LTKNG-----ATAGKATVEIWOETIAIKSVNFKPVQTFNFKKINIGTVLELEKSNLDD 725
DB 1409 LNTINGLNIISNGRNTVRLRGKIDVKYIQGVASVEEVIEAKRVLEKVKDLSDEBERET 1468
QY 726 IVK-GIN-----LTKTOHKVRVVKSG--AEQGLYLDENGDAVFNAGD 766
DB 1469 LAKLGSVAVRFPVFNNAITVNTQNEFTKPSQVITSECK-----ACFSSGN 1515
RESULT 10
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-10
Query Match 5.9%; Score 249.5; DB 2; Length 1600;
Best Local Similarity 22.2%; Pred. No. 9.6e-10;
Matches 185; Conservative 121; Mismatches 346; Indels 183; Gaps 37;
QY 16 GSINLVKGAITGKPDGTGYPTE-----SIDRASAIVFTKILNLPV 58
DB 850 GNIN-ITKANVTIQAQTSNSNTGKKRTILGNISVEGNSLITCANIVGN--LSIAE 906
QY 59 DENAQSPFKDAKNI---WSSKYIAAEKAGVGVGKNGKFNPEKIDRASPASMLVSAYN 115
DB 907 DSTFKGEASDLNITGTTNNGTANINIKGVVK-----LGDINNKGGLNITNASG 957

QY 116 L-KDKVNGELVTFPDLDDHMGEEKANILILGIVSGTGKWEPNKSVSRAAAFIALT 174
DB 958 TQKTIINGNITWEKGD--NIKNIKADAEIQIG---GNISQEGNLTISSDKVNTNQIT 1012
QY 175 DK---KYGKQNAQAVTVDKVSFETKLTLTGTGLDKLSADDVTELEGKAVAIEASTDGT 231
DB 1013 IKAGVEGGSDESSEA-----ENANLTIQTKEL-KL-AGDLNITSGFNKABITAKNGSD 1062
QY 232 SAVVTILGGKVAENKOLTV-KVKNQSFVT-----KPVYEVKKLAVEKLTFFDDRAQQAIAF 285
DB 1063 LIIGNASGNAADAKKVPDKVKDSKISTDGHNVTLNSEKTSNGSSNAGDNSTGLTISA 1122
QY 286 KUNDEKGNADVLYNLNANHDVFNANLDGSPANIF--EGGEATSTTKLAVIGIKQGDYK 343
DB 1123 K-----DV---TVNNNVTSKHTINISAAAGNVTTKEGTTINATTGTSVEVTAQNG-- 1168
QY 344 VEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNTGSKLSGKDFALN 403
DB 1169 -----TIKNITSONVTVTATENLVTTENAVINA-----TSQVNTSTKTDGDIKGGIE 1216
QY 404 SONLVVEKASLNKL-VATIGEDKVPDPGSIKSNHGIISVNVNITITAAAGAEATLT 462
DB 1217 STSGNVNITASGNITKVSNIITGQDVTV-----TADAGALTTTAGSTISATTGNANIT 1268
QY 463 IKVGQVTKDVK-----FKVTTDSRKLVSVKANPDKLOVQVQNTLPVT--FV 506
DB 1269 TKTGDIINGKVESSSGSVTLVATGATLAVNISGNTVTITADSGKLTSTVGSTINGTNSVT 1328
QY 507 TTDQYGDPPFGANTAATKEVLPKGT-----VVAEGGLDVVTTDSGSIQTKTIGVTGN 557
DB 1329 TSSQSGDIEGTISGNTVNTASTGLTIGNSAKVEAKGAATLTAESGKLTQ---TGS 1384
QY 558 DVGEQTVHFQNGGATLSLYNVNTEGNVAFKQFVLVSKVGQYQSPDKLDLNVST-TV 616
DB 1385 SI-----TSSNGQTTLTAKDSSAGNINAANVTL-NTTGLTITTGOSKINATSGTLTI 1436
QY 617 EYOLSKY-----TSRVYSDPENLEGVEYESKNLAVADAKIVGNKVVTGKTPGVKVDH-- 670
DB 1437 NAKADLGAASGDRVTVVNAASG-----SGN-----VIAKTSVVNTG 1478
QY 671 -LTKNG-----ATAGKATVEIWOETIAIKSVNFKPVQTFNFKKINIGTVLELEKSNLDD 725
DB 1479 LNTINGLNIISNGRNTVRLRGKIDVKYIQGVASVEEVIEAKRVLEKVKDLSDEBERET 1538
QY 726 IVK-GIN-----LTKTOHKVRVVKSG--AEQGLYLDENGDAVFNAGD 766
DB 1539 LAKLGSVAVRFPVFNNAITVNTQNEFTKPSQVITSECK-----ACFSSGN 1585

RESULT 11

US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match 5.9%; Score 249.5; DB 3; Length 2353;
Best Local Similarity 22.0%; Pred. No. 1.7e-09;
Matches 201; Conservative 107; Mismatches 335; Indels 269; Gaps 45;

QY 23 DKGAITGPDGTGPTSIDRASA--IFTKILNLPVDENAPQSPFKAKNIWSSKYIAA 80
DB 705 DRGKVTVK-DAT---ANDADKKVATKDVATAI-----NSAATFVKTNLTTS----- 748
QY 81 VEKAGVVGKDGKGFYPEGKIDRASAVMLV--SAYNLKDKVNGELVTTFFEDLDHWGEE 138
DB 749 -----IDENPTDNGKODALKAGDTLTFKAGKNLKVKEDGKNITF--DLAKNLEVK 797
QY 139 KANI--LINLGSVGTGGKWEPNK--SVSRABAAQFIATDKYKGNKNAQAYTVDKVS 194
DB 798 TAKVSDTLTIGNTPTGTTATPKVNTSTADGLNFAKETADASGSKN---VYLKGI--- 851
QY 195 EPTKLTLTGTGLDLSADDVTLLEGDK---AVAIBASTDGTSAVVTLGKVAAPNKDLTVKV 251
DB 852 -ATTLTSPSAGA-KSSHVDLNVDAATKKSNAASIE-----DVLKAGNIOQNGN----- 897
QY 252 KQSFVTKFVYEVKLAVEKLTFFDDRAGQAIAPKLNDEKGNADVEYLNLANHVDKVFVAN 311
DB 898 -NVDYVATY-----DTVNFTDDSTGTTTV-----TVTQ 924
QY 312 NLDGSPANIFEGGEAT---STTGKLAAGIKQGDYKVEVQVTKRGGLTVS-----NTGI 361
DB 925 KADGKADVKGAKTSVTKHNGKLTGKOLK-----ANNGATVSEDDGKDTGTGL 976
QY 362 ITVKNL-----DTPASAIKNVVFALDADNDGVNNGSKLGGK----- 399
DB 977 VTAKTVIDAVNKSQWRTGEGATAETGATV-----NAGNAETVTSGTSVNFQNGNAT 1029
QY 400 -FALNSQNLVGEKASLN-----KLVA-----TIAGEDKVVDPGSIKSSN-- 440
DB 1030 TATVSKONGNINVKYDNNVNGGLKIGDDKIVADTTTLTVTGKVSVPAGANSVNNKKL 1089
QY 441 ---HGLISVNN-----YITAEAAAGEATLTKVGDVTKVKKFVTTDSRLKVSVK- 487
DB 1090 VNAEGLATALNLSWTAADKADYADGESEGETDQEVKAGD-----KVTFKAGKNLKVQ 1142
QY 488 ANPKLQVQVQNTLPVFTVTTDQVDPFGANTAAIKVLPKGTG-VVAEGGLDVTVDSSG 546
DB 1143 SEKQFTVSLQDTLGLTSLITLG--GTANGRN-----DTGTVINKDGTLTTLANGAA 1191
QY 547 IGT-----KITGVTDNDVGEVTHFQNGNGATLGLSVNV-----TEGNVAPKNPE 592
DB 1192 AGTASNGNTISVTKDGISAGNKEITNVKSL--KTYKDTQNTADETQDKEFHAQVQAN 1249
QY 593 LVSKVQVQSQPDKLDLNVSTTVYEQ-----LSKYSRDRVSDPENLEGYEVESK 643
DB 1250 EVEFVGKNGATVSAKTDNNNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKVNDTQD-----N 1304
QY 644 NLAADAKIVGNKVVVTGKTPGKVDIHLTKNGATAGATVEIIVQETIAIKSVN----- 696
DB 1305 NLLTVDA-----TKGASVAKGEFNAVTTDTAQTGNTANERKGVVVGSGNGATATET 1356
QY 697 -FKPVQNTENFVEKIN-IGTVLELE-----KSNLDP-----IVKGINLTKETQHKVR 741
DB 1357 DKKKVATVGDVAKAINDAAATFVKVENDSDSATIDDSPTDGDANDALKAGDTLTKAGKNLK 1416
QY 742 V-----VKSAEOKGLYDRNGDAVFNAGDVK-LGDTVTSQTSDSA--- 781
DB 1417 VKRDKGNITFALANDLSVKSATVSDKLSLGTNGNKNVITSDTKGLNFAKDKSRTGDDANTH 1476
QY 782 LPNFKADLYDTL 793
DB 1477 LNGIASLTDLT 1488

RESULT 12

US-08-913-942-4

; Sequence 4, Application US/08913942

; Patent No. 6200578

GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-942-4

Query Match 5.9%; Score 249.5; DB 3; Length 2353;
Best Local Similarity 22.0%; Pred. No. 1.7e-09;
Matches 201; Conservative 107; Mismatches 335; Indels 269; Gaps 45;

QY 23 DKGAITGPDGTGPTSIDRASA--IFTKILNLPVDENAPQSPFKAKNIWSSKYIAA 80
DB 705 DRGKVTVK-DAT---ANDADKKVATKDVATAI-----NSAATFVKTNLTTS----- 748
QY 81 VEKAGVVGKDGKGFYPEGKIDRASAVMLV--SAYNLKDKVNGELVTTFFEDLDHWGEE 138
DB 749 -----IDENPTDNGKODALKAGDTLTFKAGKNLKVKEDGKNITF--DLAKNLEVK 797
QY 139 KANI--LINLGSVGTGGKWEPNK--SVSRABAAQFIATDKYKGNKNAQAYTVDKVS 194
DB 798 TAKVSDTLTIGNTPTGTTATPKVNTSTADGLNFAKETADASGSKN---VYLKGI--- 851
QY 195 EPTKLTLTGTGLDLSADDVTLLEGDK---AVAIBASTDGTSAVVTLGKVAAPNKDLTVKV 251
DB 852 -ATTLTSPSAGA-KSSHVDLNVDAATKKSNAASIE-----DVLKAGNIOQNGN----- 897
QY 252 KQSFVTKFVYEVKLAVEKLTFFDDRAGQAIAPKLNDEKGNADVEYLNLANHVDKVFVAN 311
DB 898 -NVDYVATY-----DTVNFTDDSTGTTTV-----TVTQ 924
QY 312 NLDGSPANIFEGGEAT---STTGKLAAGIKQGDYKVEVQVTKRGGLTVS-----NTGI 361
DB 925 KADGKADVKGAKTSVTKHNGKLTGKOLK-----ANNGATVSEDDGKDTGTGL 976
QY 362 ITVKNL-----DTPASAIKNVVFALDADNDGVNNGSKLGGK----- 399

Db 977 VTAKTVIDAVNKGMRVTEGATAETGATAV-----NAGNAETVTSVNFKNGNAT 1029
QY 400 -PALNSONLVGKASLN-----KLVA-----TIAGEDKVDPGSGISIKSSN-- 440
Db 1030 TATVSKDNGNINVKYDVNVDGDLKIGDDKIVADTTTLTVTGKVSVPAGANSVNNKKL 1089
QY 441 ---HGIISVNN-----YITABAAGEATLTIKVGVDVKDKVKTTSRKLVSVK- 487
Db 1090 VNAEGLATALNLSWTAADKYADGESEGETDQEVKAGD-----KVTFFKAGKLNKVKQ 1142
QY 488 ANPDKLVQVQNTLPVTFTVTDQYDGPFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS 546
Db 1143 SEKDFYSLQDTLTGLTSLTGL--GTANGRN-----DTGTVINKDGLTITLANGAA 1191
QY 547 IGT-----KTIGVTGNDVGEVTHFQNGGATLGSLYVNV-----TEGNVAFKNFE 592
Db 1192 AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNNAN 1249
QY 593 LYSKVQYQSGSPDKLDLNVSTTVEYQ-----LSKYTSDRVYSDPENLGEYEVESK 643
Db 1250 EVEFVGKNGATVSAKTDNNNGKHTVTDVABAKVGDGLEKDTGKILKLVNDTGD-----N 1304
QY 644 NLAVADAKIVGNKVVVTGTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVN----- 696
Db 1305 NLLTVDA-----TKGASVAKGEFNAVTTDTAAGTNANERGVVVKSGNGATATET 1356
QY 697 -FKPVQTEFVEKKIN-IGTVLELE---KSNLDD-----IVKGINLTKEHQKVR 741
Db 1357 DRKKVATVGDVAKAINDAATFVKVENDSATIDDSPTDDGANDALKAGDGLTLKAGKLNK 1416
QY 742 V-----VKSAGAEQKGLYLDNRGDAVFNAGDVK-LGDVTVTSQTSDSA--- 781
Db 1417 VKRDGKNITFALANDLSVKSATVSDKLSLGTNGKNVNTSDTKGLNFAKDSKTGDDANIH 1476
QY 782 LPNFKADLYDTL 793
Db 1477 LNGIASTLTDTL 1488

RESULT 13

US-09-669-974-33
; Sequence 33, Application US/09669974
; Patent No. 633173
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/669,974
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-669-974-33

Query Match 5.9%; Score 249.5; DB 4; Length 2353;
Best Local Similarity 22.0%; Pred. No. 1.7e-09;
Matches 201; Conservative 107; Mismatches 335; Indels 269; Gaps 45;
QY 23 DKGAITGKDPGYGPTESIDRASAAY--IFTKILNLPVDENAPQSFKAONKNSKYTAA 80
Db 705 DRGKVTVK-DAT---ANDADKKVATYKDVATAI-----NSAATFVKTNLTTS----- 748

RESULT 14

US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. 6607729
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03

QY 81 VEXGVVVGKDGKGFYFEGKIDRASFAFMLV--SAYNLKDKVNGELVTTFTEDLLDHGEE 138
Db 749 -----IDENPTDNGKDDALKAGDGLTLTFKAGKLNKVKRDGKNITP--DLAKNLEVK 797
QY 139 KANI--LINLIGISVGTGGKWEPNK--SVSPAEEAQAIFALTDDKYYKKDKNAQAAVTVDKVS 194
Db 798 TAKVSDTLTITGGNTPTGGTTATPKVNITSTADGLNFAKETADASGSN---VYLKGI--- 851
QY 195 EPTKLTITGTGLDKLADDDVTLLEGDK---AVAIEASTDGTSAVVTILGGKVPAPNKDLTVKV 251
Db 852 -ATLLEPSAGA-KSSHVDLNVDAATKKSNAASIE-----DVLRAAGNIQNGN----- 897
QY 252 KNSQSVTKFYVYVKLAVEKLTPTDDRAGQAIAFLKNDKGNADVEYLNLANHNDVAFVAN 311
Db 898 -NDVYVATY-----DTVNFDDSTGTTTV-----TVTQ 924
QY 312 NLDGSPANIPEGGEAT---STTGKLAVGIKQGDYKVEVQVTKEGGLTVS-----NTGI 361
Db 925 KADGKADGVKIGAKTSVIKDHNGKLTFTGKDLK-----ANNGATVSEDDCKDGTGTL 976
QY 362 ITVKNL-----DTPASAIKNVVFALDADNDGVNVYSGKLSGKD----- 399
Db 977 VTAKTVIDAVNKGMRVTEGATAETGATAV-----NAGNAETVTSVNFKNGNAT 1029
QY 400 -PALNSONLVGKASLN-----KLVA-----TIAGEDKVDPGSGISIKSSN-- 440
Db 1030 TATVSKDNGNINVKYDVNVDGDLKIGDDKIVADTTTLTVTGKVSVPAGANSVNNKKL 1089
QY 441 ---HGIISVNN-----YITABAAGEATLTIKVGVDVKDKVKTTSRKLVSVK- 487
Db 1090 VNAEGLATALNLSWTAADKYADGESEGETDQEVKAGD-----KVTFFKAGKLNKVKQ 1142
QY 488 ANPDKLVQVQNTLPVTFTVTDQYDGPFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS 546
Db 1143 SEKDFYSLQDTLTGLTSLTGL--GTANGRN-----DTGTVINKDGLTITLANGAA 1191
QY 547 IGT-----KTIGVTGNDVGEVTHFQNGGATLGSLYVNV-----TEGNVAFKNFE 592
Db 1192 AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNNAN 1249
QY 593 LYSKVQYQSGSPDKLDLNVSTTVEYQ-----LSKYTSDRVYSDPENLGEYEVESK 643
Db 1250 EVEFVGKNGATVSAKTDNNNGKHTVTDVABAKVGDGLEKDTGKILKLVNDTGD-----N 1304
QY 644 NLAVADAKIVGNKVVVTGTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVN----- 696
Db 1305 NLLTVDA-----TKGASVAKGEFNAVTTDTAAGTNANERGVVVKSGNGATATET 1356
QY 697 -FKPVQTEFVEKKIN-IGTVLELE---KSNLDD-----IVKGINLTKEHQKVR 741
Db 1357 DRKKVATVGDVAKAINDAATFVKVENDSATIDDSPTDDGANDALKAGDGLTLKAGKLNK 1416
QY 742 V-----VKSAGAEQKGLYLDNRGDAVFNAGDVK-LGDVTVTSQTSDSA--- 781
Db 1417 VKRDGKNITFALANDLSVKSATVSDKLSLGTNGKNVNTSDTKGLNFAKDSKTGDDANIH 1476
QY 782 LPNFKADLYDTL 793
Db 1477 LNGIASTLTDTL 1488

```

; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

Query Match          5.9%; Score 249.5; DB 4; Length 2353;
Best Local Similarity 22.0%; Pred. No. 1.7e-09;
Matches 201; Conservative 107; Mismatches 335; Indels 269; Gaps 45;

QY 23 DKGAITGPDTGYPTESIDRASA--IFTKILNLPVDENAPSFKDAKNIWSSKIIAA 80
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 705 DRGKVTVK-DAT---ANDADKKVATVDVATAI-----NSAATFVKTENLTTS--- 748
QY 81 VEKAGVVGDKGFENYFEGKIDRASFAFMLV--SAYNLKDKUNGELVTTFFEDLLDHGEE 138
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 749 -----IDEDNPTDNGKODALKAGDTLTFKAGKNLKVCRDGNKITF--DLAKNLEVK 797
QY 139 KANI--LINLGISVGTGGKWEPNK--SVSRAEAAQFIALTDKKYGKDNAAQAVVTDVKVS 194
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 798 TAKVSDTLTIGNTPTGGTATPKVNITSTADGLNFAKETADASGSKN---VYLKGI--- 851
QY 195 EPTKLTITGTGLDKLSADDDVLEGDK---AVAIEASTDGTSAVVTGGKVPAPKDLTVKV 251
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 852 -ATLTPEPSAGA-KSSHVDLNVDAATKKSNAASIE-----DVLRAGWNIOGNGN---- 897
QY 252 KQSFVTKFVYEVKLAVEKLTFPDDRAGQAIAPFLNDEKGNADVEYLNLANHDKVFVAN 311
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 898 -NVDYVATY-----DTVNFDDSTGTTV-----TVTQ 924
QY 312 NLDGSPANIFEGGEAT---STTGKLVAGIKOGDYKVEVQVTKRGGLTVS-----NTGI 361
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 925 KADGKGADVTKIGAKTSVIKDHNGKLFQTKDLK-----ANNGATVSEDDGKDTGTGL 976
QY 362 ITVKNL-----DTPASAIKNVVFALDADNGVNVYSGKLSGKD-----399
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 977 VTAKTVIDAVNKSQWRVTGEGATETGATAV-----NAGNAETVTSGETSVNFKNAT 1029
QY 400 -FALNSQNLVGEKASLN-----KLVA-----TIAGEDKVPDPSISIKSN-- 440
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1030 TATVSKDNGNINVKYDVNVGDLKIGDDKKIVADTTTLTVTGKVSVPAGANSVNNKKL 1089
QY 441 ---HGIISVNN-----YITAEAAAGEATLTIKVGDVTKDKVFKVTTDSRKLVSVK- 487
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1090 VNAEGLATALNLSWTAKADKYADGESEGETDQEVKAGD-----KVTFAKGNLKVQK 1142
QY 488 ANPKQLQVQNKTLFVTFVTTDQYGDPPFGANTAAIKEVLPTKG--VVAEGGLDVVTTDSGS 546
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1143 SEKDFTYSLODITLGLTSITLG--GTANGRN-----DTGTVINKDGLTITLANGAA 1191
QY 547 IGT-----KTIGVTGNDVGEVTVHFQNGATLGSLLVNV-----TEGNAFKNFE 592
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1192 AGTDSNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFFAAVKAN 1249
QY 593 LVSKVGQVQSPDKLDLNVSTTVEYO-----LSKVTSDRVYSDPENLSEGYVESK 643
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1250 EVEFVKNKATVSAKTDNNGHVTVIDVAEKVGDGLEKDTGKIKLVNDTGG-----N 1304
QY 644 NLAVADAKIVGNKVVVTGTPGVVDIHLTKNGATAGATVETVQETIAIKSVN-----696
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1305 NLLTVDA-----TKGASVAKGEFNAVTTTDAQTGNTANERGVVVKVSGSGATATET 1356
QY 697 -FKPVQTEFVEKKIN-IGTVLELE---KSNLDD-----IVKGINLTKETQHKVR 741
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1357 DKKKVATVGDVAKALNDAAATFVRVNDSDATIDSDPTDGDANDALKAGDTLTLAKGNL 1416
QY 742 V-----VKGAEQGLYLDRNGDAVFNAGDVK-LGDVTVVSQTSDSA---781

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:29:55 ; Search time 18 seconds

(without alignments)

2409.691 Million cell updates/sec

Title: US-09-844-281-1

Perfect score: 4202

Sequence: 1 AGKSFDPVAGHWAEGSINY.....ITSEIGSQAVHVNLENPL 833

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	4202	100.0	862	1	SLA2_BACAA
2	2833.5	67.4	874	1	SLAP_BACLI
3	719	17.1	814	1	SLAI_BACAA
4	529	12.6	531	1	YO42_BACAN
5	348	8.3	1176	1	SLAP_BACSH
6	265.5	6.3	762	1	SLAP_ACEKI
7	232.5	5.0	941	1	GUN_FACSS
8	230	5.5	1325	1	YDEK_ECOLI
9	223.5	5.3	2003	1	YDBA_ECOLI
10	219	5.2	1116	1	SLPH_BRECH
11	217	5.2	1655	1	OLPM_RICCN
12	211	5.0	1053	1	SLPM_BACBR
13	207.5	4.9	1645	1	OLPM_RICTY
14	207.5	4.9	2358	1	YEEJ_ECOLI
15	207.5	4.9	2660	1	YEEJ_ECO57
16	206.5	4.9	1654	1	OLPM_RICRI
17	206	4.9	1698	1	41_DROME
18	200	4.8	1643	1	OLPM_RICPR
19	199.5	4.7	939	1	SLAP_CAMFE
20	197.5	4.7	1608	1	HLVA_SERMA
21	195.5	4.7	1848	1	CBPA_CLOCL
22	194.5	4.6	1656	1	OLPM_RICJA
23	194.5	4.6	1902	1	P2P_EACPA
24	194	4.6	1300	1	120K_RICRI
25	193	4.6	917	1	SLAP_THETH
26	189.5	4.5	1183	1	CNA_STAAU
27	189	4.5	1953	1	BIGA_SALTY
28	188.5	4.5	1276	1	PMP6_CHLPN
29	187.5	4.5	1861	1	APU_THETU
30	185.5	4.4	1577	1	HLVA_PROMI
31	183.5	4.4	1087	1	XYNK_CLOTM
32	183	4.4	2249	1	OLPM_RICRI
33	183	4.4	4705	1	FAT2_DROME

34 181.5 4.3 1228 1 SLAP_BACST
35 181.5 4.3 1902 1 P2P_LACLC
36 180 4.3 1286 1 AIDA_ECOLI
37 180 4.3 1569 1 YFUA_ECOLI
38 179.5 4.3 1726 1 MSP1_PLAFC
39 179.5 4.3 1726 1 MSP1_PLAFC
40 179 4.3 1664 1 SLP1_CLOTM
41 178.5 4.2 2021 1 OMPA_RICCN
42 176.5 4.2 447 1 ANCA_CLOTM
43 176.5 4.2 1902 1 PIP_LACLC
44 174 4.1 1025 1 SLAP_CAUCR
45 172 4.1 1020 1 BCA_STRAG

ALIGNMENTS

RESULT 1
ID SLA2_BACAA STANDARD; PRT; 862 AA.
AC P34217;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE S-layer protein EAL precursor.
GN EAG OR BA0887.
OS Bacillus anthracis (strain Ames), and
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=198094, 1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames;
RX MEDLINE=22608414; PubMed=13721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Heigason E., Ristone J., Wu M.,
RA Kolonay J.F., Seaman M.O., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance J.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Clime R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.,
RT The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.
RL Nature 423:81-86 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sterne / 9131;
RX MEDLINE=97260111; PubMed=9106206;
RA Mesnage S., Tsigi-Couture E., Mock M., Gounon P., Fouet A.,
RT Molecular characterization of the Bacillus anthracis main S-layer
RT component: evidence that it is the major cell-associated antigen.
RL Mol. Microbiol. 23:1147-1155 (1997).
CC -1- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC of proteins which coat the surface of bacteria.
CC -1- SUBCELLULAR LOCATION: Cell wall.
CC -1- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

McSnay et al

EMBL: AS017027; RAP24884.1;
DR ENBL; X99724; CAA68063.1;
DR TIGR; BA0887;
DR InterPro; IPR001119; SLH.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:29:55 ; Search time 18 seconds
(without alignments)
2409.691 Million cell updates/sec

Title: US-09-844-281-1

Perfect score: 4202

Sequence: 1 AKSPDPVAGHWAEGSYN.....ITSEIGSOAVHVNLPNL 833

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4202	100.0	862	1 SLA2_BACAA	P94217 bacillus an
2	2833.5	67.4	874	1 SLAP_BACLI	P49052 bacillus li
3	219	17.1	814	1 SLA2_BACAA	P49051 bacillus an
4	529	12.6	531	1 YQ42_BACAN	Q9RM20 bacillus an
5	348	8.3	1176	1 SLAP_BACSH	P38537 bacillus sp
6	265.5	6.3	762	1 SLAP_ACEKI	P22528 acetogenium
7	252.5	6.0	941	1 GUN_EACSE	P19424 bacillus sp
8	230	5.5	1325	1 YBK_ECOLI	P32051 escherichia
9	223.5	5.3	2003	1 YBK_ECOLI	P33666 escherichia
10	219	5.2	1116	1 YBK_ECOLI	P38538 brevivibacill
11	217	5.2	1655	1 OMPB_RICCN	Q9KKA3 r outer mem
12	211	5.0	1053	1 OMPB_RICCN	P06546 bacillus br
13	207.5	4.9	1645	1 OMPB_RICCN	P06546 bacillus br
14	207.5	4.9	2358	1 YBK_ECOLI	P38537 escherichia
15	207.5	4.9	2660	1 YBK_ECOLI	P38537 escherichia
16	206.5	4.9	1654	1 OMPB_RICCN	Q9KKA3 r outer mem
17	206	4.9	1698	1 41_DROME	Q9V8R9 drosophila
18	200	4.8	1643	1 OMPB_RICCN	Q9KKA3 r outer mem
19	199.5	4.7	939	1 SLAP_CAMPF	P35020 r outer mem
20	197.5	4.7	1608	1 HLYA_SERMA	P35827 campylobact
21	195.5	4.7	1848	1 CPBA_CLOCL	P5320 serratia ma
22	194.5	4.6	1656	1 OMPB_RICCN	P38058 clostridium
23	194.5	4.6	1902	1 P2P_LACPA	Q06653 r outer mem
24	194	4.6	1300	1 P2P_LACPA	Q06653 r outer mem
25	193	4.6	1917	1 120K_RICRI	Q2470 lactobacill
26	189.5	4.5	1183	1 SLAP_THETH	P14914 rickettsia
27	189	4.5	1953	1 CNA_STAAU	P35830 thermus the
28	188.5	4.5	1276	1 BIC6_SALTY	Q3654 staphylococ
29	187.5	4.5	1861	1 PWE6_CHLPN	Q38927 salmonella
30	185.5	4.5	1577	1 YBK_ECOLI	Q38927 salmonella
31	185.5	4.4	1087	1 HLYA_PROMI	P36466 proteus mir
32	183	4.4	2249	1 OMPA_RICCN	P38535 clostridium
33	183	4.4	4705	1 FAT2_DROME	P15921 rickettsia

34	181.5	4.3	1228	1	SLAP_BACST	P35825 bacillus st
35	181.5	4.3	1902	1	P2P_LACLC	P15293 lactococcus
36	180	4.3	1286	1	AIDA_ECOLI	Q03155 escherichia
37	180	4.3	1569	1	YFJA_ECOLI	P52143 escherichia
38	179.5	4.3	1726	1	MSPI_PLAFC	P04934 plasmodium
39	179.5	4.3	1726	1	MSPI_PLAFC	P50495 plasmodium
40	179	4.3	1664	1	SLPI_CLOTM	Q06852 clostridium
41	178.5	4.2	2021	1	OMPA_RICCN	Q52857 rickettsia
42	176.5	4.2	447	1	ANCA_RICCN	Q06848 clostridium
43	176.5	4.2	1902	1	P1P_LACLC	P16271 lactococcus
44	174	4.1	1025	1	SLAP_CAUCR	P35828 caulobacter
45	172	4.1	1020	1	BCA_STRAG	Q02192 streptococc

ALIGNMENTS

RESULT 1

SLA2_BACAA STANDARD; PRT; 862 AA.

AC P94217;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE S-layer protein EAI precursor.

GN EAG OR BA0887.

OS Bacillus anthracis (strain Ames), and

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OC NCBI_TaxID=198094, 1392;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Ames;

RC MEDLINE=22608414; PubMed=12721629;

RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T., Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R., Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M., Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gwinn M., DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H., Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D., Berton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F., Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Niernan W.C., Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B., Fraser C.M.;

RT "The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria";

RL Nature 423:81-86 (2003).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=sterne / 9131;

RC MEDLINE=97260111; PubMed=9106206;

RA Mesnage S., Tosi-Couture E., Mock M., Gounon P., Fouet A.;

RT "Molecular characterization of the Bacillus anthracis main S-layer component: evidence that it is the major cell-associated antigen";

RL Mol. Microbiol. 23:1147-1155 (1997).

CC -1- FUNCTION: The S-layer is a paracrystalline mono-layered assembly of proteins which coat the surface of bacteria.

CC -1- SUBCELLULAR LOCATION: Cell wall.

CC -1- SIMILARITY: Contains 3 S-layer homology (SLH) domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

DR EMBL; AB017027; AAP24884.1; ..

DR EMBL; X99724; CAA68063.1; ..

DR TIGR; BA0887; ..

DR InterPro; IPR001119; SLH.

Mesnage et al

```
DR Pfam; PF00395; SLH; 3.
KW Signal; Repeat; Cell wall; S-layer; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 862 S-LAYER PROTEIN EAL.
FT DOMAIN 34 76 SLH 1.
FT DOMAIN 95 136 SLH 2.
FT DOMAIN 157 197 SLH 3.
SQ SEQUENCE 862 AA; 91362 MW; CB169202F62CCCA0 CRC64;

Query Match 100.0%; Score 4202; DB 1; Length 862;
Best Local Similarity 100.0%; Pred. No. 1.3e-175;
Matches 833; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGKSPDPVAGHWAEGSINLVKGAITGKPDGTGPTESIDRASAAVFTKILNLPVDE 60
DB 30 AGKSPDPVAGHWAEGSINLVKGAITGKPDGTGPTESIDRASAAVFTKILNLPVDE 89
QY 61 NAQSPFKDANKINWSKYIAAVEKAGVVGKDGKENTYPEGKIDRASFAFMLVSAYNLKDV 120
DB 90 NAQSPFKDANKINWSKYIAAVEKAGVVGKDGKENTYPEGKIDRASFAFMLVSAYNLKDV 149
QY 121 NGELVTFPEDLLDHGEEKANILINLIGSVGTGKWEPNKSVSRAEAAQFIALTDKKYG 180
DB 150 NGELVTFPEDLLDHGEEKANILINLIGSVGTGKWEPNKSVSRAEAAQFIALTDKKYG 209
QY 181 KDNAQAVYVDVKVSEPTKLTGTGLDKLSADDVTLEGDKVAIEASTDGTSAVVTILGGK 240
DB 210 KDNAQAVYVDVKVSEPTKLTGTGLDKLSADDVTLEGDKVAIEASTDGTSAVVTILGGK 269
QY 241 VAPNKDLTVKVNQSVFTKVFYVEVKLAVEKLTDDDRAGQAIAFKLNDEKGNADVEYL 300
DB 270 VAPNKDLTVKVNQSVFTKVFYVEVKLAVEKLTDDDRAGQAIAFKLNDEKGNADVEYL 329
QY 301 LANHDKVFAVNNLDGSPANIPEGGEATSTTGKLAIVGKQDYKVEVQVTKRGGLTVSNTG 360
DB 330 LANHDKVFAVNNLDGSPANIPEGGEATSTTGKLAIVGKQDYKVEVQVTKRGGLTVSNTG 389
QY 361 IITVKNLTPASAIKNVVFALDADNDGVNYSKLSGKDFALNSQNLVGEKASLNKLV 420
DB 390 IITVKNLTPASAIKNVVFALDADNDGVNYSKLSGKDFALNSQNLVGEKASLNKLV 449
QY 421 TIAGEDKVVDPGSIKSNHGIIISVANNYITAEAGAEATITIKVGDVTKDVKFKVTTDS 480
DB 450 TIAGEDKVVDPGSIKSNHGIIISVANNYITAEAGAEATITIKVGDVTKDVKFKVTTDS 509
QY 481 RKLVSFKANPDKLQVQVQNTLPTVFTVTTDQGDPPGANTAAIKVLPKTVGVAEGGLDV 540
DB 510 RKLVSFKANPDKLQVQVQNTLPTVFTVTTDQGDPPGANTAAIKVLPKTVGVAEGGLDV 569
QY 541 TTDSGSIKTLTGVTGNDVGEHTVFONGGATIGSLVNVNTEGNAVAFKNFELSVKVGQY 600
DB 570 TTDSGSIKTLTGVTGNDVGEHTVFONGGATIGSLVNVNTEGNAVAFKNFELSVKVGQY 629
QY 601 GQSPDKLDLVSTTVEYQSKYTSRVSVDPENLEGYVESKQNLAVADAKIVGNKVVT 660
DB 630 GQSPDKLDLVSTTVEYQSKYTSRVSVDPENLEGYVESKQNLAVADAKIVGNKVVT 689
QY 661 GKTGPGVDIHLTKNGATAGATVEIYQETIAIKSVNEKPVQTEFVEKINIGTVLELEK 720
DB 690 GKTGPGVDIHLTKNGATAGATVEIYQETIAIKSVNEKPVQTEFVEKINIGTVLELEK 749
QY 721 SNLDDIVKGINLTQKHQKVRVYKSGAEQKLYLDRNGDAVFNAGDVKLGDVTVTSQSDS 780
DB 750 SNLDDIVKGINLTQKHQKVRVYKSGAEQKLYLDRNGDAVFNAGDVKLGDVTVTSQSDS 809
QY 781 ALPNFKADLYDTLTTKYTDKGLTVFKVKDKQVITSEIGSQAVHVNVLNPNL 833
DB 810 ALPNFKADLYDTLTTKYTDKGLTVFKVKDKQVITSEIGSQAVHVNVLNPNL 862

RESULT 2
P1: us-09-844-281-1.1.rsp
P2: BACLI
P3: SLAP_BACLI
P4: STANDARD; PRT; 874 AA.
```